Cell-Free Enabled Bioproduction and Biological Discovery

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

David Cinay Garcia

May 2020

Copyright © 2020 by David Garcia.

All rights reserved.

DEDICATION

This dissertation is dedicated to all the people that supported me throughout my education.

To my parents, Abel Garcia and Reyna Garcia for giving me all the love and support I could have ever needed.

And to my many instructors and mentors for their guidance and patience.

Esta disertación está dedicada a todas las personas que me apoyaron durante mi educación.

A mis padres, Abel García y Reyna García, por dando me todo el amor y el apoyo que podría haber necesitado.

I a mis muchos instructores y mentores por su orientación y paciencia.

ACKNOWLEDGEMENTS

I would like to acknowledge all of the support given to me by my friends and family and to the co-authors whose efforts contributed to this work. I would also like to thank Dr. Amber B. Webb, Carmen Foster, and Dr. Robert Standaert for sharing their experience, technical expertise, and patience with me these last 5 years. I am grateful to the Bredesen Center, Oak Ridge National Laboratory's talented scientists, and to my advisor Mitch Doktycz for inviting me into a wonderful institution at which to complete a doctorate. And I am grateful to Katie Warczak and Po Archie for always being there when I needed them.

Abstract

As our understanding of the microbial world has progressed, so too has the backlog of information and open questions generated from the thousands of uncharacterized proteins and metabolites with potential applications as biofuels, therapeutics, and biomaterials. To address this problem, new tools need to be developed in order to rapidly test and take advantage of uncharacterized proteins and metabolites. Cell-free systems have developed into a high-throughput and scalable tool for synthetic biology and metabolic engineering with applications across multiple disciplines. The work presented in this dissertation leverages cell-free systems as a conduit for the exploration of protein function and metabolite production using two complementary approaches. The first elucidates interaction networks associated with secondary metabolite production using a computationally assisted pathway description pipeline that employs bioinformatic searches of genome databases, structural modeling, and ligand-docking simulations to predict the gene products most likely to be involved in a metabolic pathway. In vitro reconstructions of the pathway are then modularly assembled and chemically verified in *Escherichia coli* lysates in order to differentiate between active and inactive pathways. The second takes a systems and synthetic biology approach to engineer *Escherichia coli* extracts capable of directing flux towards specific metabolites. Using growth and genome engineering-based methods, we produced cell-free proteomes capable of creating unconventional metabolic states with minimal impact on the cell in vivo. As a result of this work, we have significantly

expanded our ability to use cell extracts outside of their native context to solve metabolic engineering problems and provide engineers new tools that can rapidly explore the functions of proteins and test novel metabolic pathways.

Table of Contents

Abstract v
Table of Contents vii
List of Tables ix
List of Figures x
1. Introduction 1
History of Cell-Free Systems 2
Scope of Dissertation 4
Cell-Free Enabled Biological Discovery 4
Engineering Cell-Free Metabolism 6
Organization of Dissertation 8
2. Computationally Guided Discovery and Experimental Validation of Indole-3-acetic Acid Synthesis Pathways 10
Abstract 11
Introduction 12
Results and Discussion 16
Bioinformatic Analyses Indicate Complex IAA Metabolism in *Pantoea sp.* YR343 16
Computational screening of ligand-docking interactions 18
Pathway verification 20
Methods 30
Pathway and gene identification 30
Ligand-docking simulations 31
Enriched extract preparation 32
In vitro reactions 33
Analysis of IAA production 34
Appendix 35
3. Elucidating the potential of crude cell extracts for producing pyruvate from glucose 43
Abstract 44
Introduction 45

Results and Discussion 48
Effects of Growth Conditions on Proteomes of Extracts 49
Pyruvate Production 55
Conclusion 63
Methods 63
Cell-free extract preparation 63
Cell-free reactions 64
Analytical measurements 65
Proteomics 65
Statistical analysis 67
Appendix 68
4. A Lysate Proteome Engineering Strategy for enhancing Cell-Free Metabolite Production 72
Abstract 73
Introduction 74
Results and Discussion 77
Methods 89
Conclusion 95
Appendix 96
5. Conclusions 100
References 105
Vita 118

List of Tables

Table 2.1. Representatives for enzymes used in BLAST analysis.....37
Table 2.2. List of template structures used for protein homology modeling.....38
Table 2.3. Enzymes identified through a BLAST search of *Pantoea* sp. YR343's genome.....39
Table 3.4. Summary of gene set enrichment analysis. 54
Table 3.5. Percent glucose consumed and the percent glucose converted to pooled pyruvate. 56
Table 4.6. Gene and protein information for MAGE targets with a potential effect on pyruvate metabolism.....78
Table 4.7. Strains created for lysate engineering study.....81
Table 4.8. MAGE oligos use for this study.....97
Table 4.9. MASC-PCR oligos used for this study. 97

List of Figures

Figure 2.1. General metabolic model of the conversion of tryptophan to IAA. ...18
Figure 2.2. Homology modeling and ligand docking to pmi39_00059 IPDC. 19
Figure 2.3. Modular assembly of potential IAA producing pathways. 23
Figure 2.4. Potential IAA-related proteins were expressed in *E. coli* and tested for solubility. Arrows mark the presence of the resultant gel band. 36
Figure 2.5. Gene neighborhoods for each of the YR343 enzymes expressed and found to be part of an IAA reaction. 36
Figure 2.6. Representative examples of binary designation used to describe binding and non-binding ligands. 40
Figure 2.7. SDS-PAGE gel of His-Tagged IAA-related proteins expressed in *E. coli* and purified. Arrows mark the presence of the resultant gel band. 41
Figure 2.8. Assembly of purified IAA producing enzymes. 42
Figure 2.9. HPLC analysis of IAA. A. Injection of 5 µg/mL purified IAA. B. Representative cell-free reaction following injection into HPLC. 42
Figure 3.10. Venn diagram of differences in protein content from extracts produced from 4 different Growth conditions. 52
Figure 3.11. Proteomic assessment of the potential in vitro metabolic fates of glycolytic intermediates. 53
Figure 3.12. Simultaneously measurements of glucose and pyruvate at various time points over a 24-hour period. 55
Figure 3.13. Simultaneously measurements of glucose and pyruvate at various time points over a 24-hour period. 57
Figure 4.14. Graphical representation of genome and lysate engineering. 77
Figure 4.15. Glycolysis and engineered pathway nodes. The number 1, 2, 3, and 4 indicate the enzymes pflB, ldhA, ppsA, aceE. 79
Figure 4.16. Source strain multiplex genome engineering and expected metabolic phenotypes of derived lysates post-reduction. 80
Figure 4.17. Comparisons of pyruvate concentration over time in WT, 6xHis-1, 6xHis-2, 6xHis-3, and 6xHis-4 lysates. 84
Figure 4.18. Proteomic analysis of an unreduced cell-free extract (blue), and a reduced cell-free extract (orange). 88
Figure 4.19. Growth rate and terminal OD600 was measured using for the four main strains produced for this study. 97
Figure 4.20. Glucose measured over a 24-hour time period. 98
Figure 4.21. Comparison of pyruvate concentration changes when fed 100 mM glucose for an extract with a single reduction of ldhA to an extract reduced for all 4 proteins *his*-ldhA, *his*-pflB, *his*-ppsa, *his*-aceE. 98
Figure 4.22. Comparison of pyruvate consumption using reduced WT and Δ4 extracts. 99

Chapter 1

INTRODUCTION

History of Cell-Free Systems

The use of cell-free systems is a relatively old set of techniques starting with one of its first uses as a means of elucidating the incorporation of amino acids into proteins. Various publications using a variety of extract donor cells followed and culminated in the first use of an *E. coli* cell-free extract to connect nucleotide triplets and the amino acids they encoded by Nirenberg and co-workers, ultimately leading to the decryption of the genetic code[1–4]. Though the technique was used for decades as a small-scale protein production system, refined protocols developed in the last 20 years significantly improved its function and generated interest in the further development of cell-free systems[5]. Advances in cell-free expression, for instance, have dramatically improved its use as a method for the expression of membrane-bound proteins, a normally difficult task when attempted with traditional methods. Since these early efforts, the field has begun to expand the functions of cell-free systems, with one of the main demarcations being that of the basic functions of the cell-free system being geared either towards peptide production in the form of cell-free protein synthesis (CFPS) or metabolite production in the form of cell-free metabolic engineering (CFME)[5–9].

The earliest efforts in CFPS took place decades before the advent of synthetic biology with Spirin and coworkers using a cellular extract of *Escherichia coli* capable of in vitro translation from mRNA templates[10]. Shortly thereafter, coupled in vitro transcription and translation was carried out from a heterologous DNA template in the crude S30 extract of *E. coli*[11]. These early efforts led to significant

advances in the field of cell-free production systems with a major milestone reached in 2001 with the publication of Ueda's work reconstituting *E. coli's* transcription and translation machinery using entirely purified components[12]. Advances in commercial optimization and fundamental research have iterated upon Ueda's original design providing one of the most popular in vitro research tools and protein production platforms[13–15]. Higher-order functions such as membrane protein production and post-translational modifications, though difficult to replicate in *E. coli* extracts, have been generated through the use of extracts from more eukaryotic organisms such as wheat germ, insect, and CHO cells[16–18].

Though much of the research concerning cell-free systems has focused on protein synthesis, the robust nature of extracts has also shown them to be effective bioproduction tools for biofuels, therapeutics, and biomaterials. The field of cell-free metabolic engineering takes advantage of the natural function of glycolysis in the cell-free proteome to rapidly prototype metabolic pathways and to serve as a bioproduction platform in its own right. The S30 CFPS system serves as the basis of much of the early CFME work as simple heterologous pathways were expressed or exogenously introduced into S30 CFPS extracts[19,20]. The process of coupling glycolysis in crude lysates with external enzymes and the necessary cofactors enabled a significant amount of control over metabolic flux compared to living systems. Whereas regulatory mechanisms and the need to maintain cell viability would normally impair the function of an in vivo bioproduction system, the removal

of the cell-free proteome from the cellular context eliminates this need and permits biosynthetic pathways to be prototyped and engineered in vitro.

Scope of Dissertation

The projects described in this thesis have been designed to facilitate the use of cell-free systems for use in research and industry. Cell-free systems probe cellular metabolism outside of their native context providing a more accessible environment for measurements of biological functions. This is leveraged towards the development of three cell-free technologies. The first is a cell agnostic method of analyzing secondary metabolism using computational tools and cell-free modular assembly of pathway components. The second shows the potential of growth enabled engineering of cell-free metabolic engineering systems. The third, develops a technique to direct flux in cell-free metabolism through the engineering of the cell-free proteome. These advances provide the tools necessary to further the use of cell-free systems for understanding and harnessing the value of the microbial world.

Cell-Free Enabled Biological Discovery

Our continued study of the microbiological world has created a backlog of information and open questions. Large collections of genomic information like the National Center for Biotechnology Information (NCBI) house over 40M proteins alone, the overwhelming number of which remain uncharacterized[21]. The information thus leaves thousands of potentially useful proteins and metabolites

uncharacterized and unused. From a scientific point of view, this limits the amount of information we can harness when developing new studies as our current understanding of potential proteins or products can be largely theoretical. From an economic view, the resources required to study even a fraction of the backlog makes the endeavor prohibitively expensive with current methods. As a result, we are left with thousands of proteins and metabolites with potential applications as biofuels, therapeutics, and biomaterials left without study. To address these problems, new tools need to be developed in order to rapidly test and take advantage of uncharacterized proteins and metabolites. To further this effort, we are pursuing the study of cell-free systems as a conduit for the exploration of protein function and metabolite production.

Cell-free systems have developed into a powerful tool for synthetic biology and metabolic engineering with applications across multiple disciplines. Developments in cell-free system technology over the last 10 years have remarkably improved its capacity for expressing proteins as well as created the field of cell-free metabolic engineering[19,20,22,23]. Our ability to use cell extracts outside of their native context to solve metabolic engineering problems has provided an excellent opportunity to engineer new tools that can rapidly explore the function of proteins and test new metabolic pathways[19,24]. As cell-free extracts are devoid of the need to propagate, production is limited only by the efficacy of the production system. As a result, there are an almost limitless number of metabolites and peptides that can be expressed using cell-free systems.

For the first section of this work, I have developed a scalable testing platform for protein function[25]. Automated annotations while informative for high-throughput analysis, are not a perfect substitute for empirical experiments and poor annotations due to lack of characterization are common. The inability to confidently predict the function of an enzyme and its place in a metabolic pathway has become a consistent problem in the study of microbial organisms, therefore, new technologies must be developed to quickly and accurately elaborate the function of metabolic pathways and individual organisms. Through the production of uncharacterized proteins using crude extracts coupled with computational tools such as bioinformatic analysis and ligand-docking simulations, we hope to rapidly generate and test hypotheses regarding the function of individual proteins and complete pathways in the organism *Pantoea* sp. YR343.

Engineering Cell-Free Metabolism

Work performed in the laboratories of Michael Jewett and James Bowie, have sought to develop cell-free systems into easily engineered bioproduction platforms. In Jewett's case, the need to prototype metabolic pathways has become one of the driving forces behind their work[26,27]. Traditional methods of generating clones with heterologous expression plasmids are limited as the cell may not be able to produce the metabolite of interest without suffering from toxicity[5]. Cell-free systems help resolve this problem, by expressing individual proteins in a cell-free extract, thereby allowing for modular combinations of extracts to be tested for production without the need for a viable expression strain. To date, the Jewett lab

has shown that crude-cell free extracts derived from *E. coli* can be used to generate glycolytic intermediates and power the production of metabolites such as 2,3-butanediol and n-butanol[19,20]. The Bowie lab has taken a similar approach towards the production, of the bioplastic polyhydroxybutyrate[28]. Interestingly, they take an approach similar to the PURE system by employing purified enzymes to create their bioproduction system[12]. This approach allows them to maintain control over cofactor recycling and balance thus eliminating the need for the cell's regulatory networks. Both approaches take advantage of a stable set of proteins capable of functioning outside of the cell, but in environments with different levels of complexity. In the case of the purified systems, attempts to make limonene, led to a maximum productivity of ~0.1 g l^{-1} h^{-1} over a 5-day period[24]. Significantly higher rates of production were seen in the case of 2,3-butanediol (11.3 g l^{-1} h^{-1}) over a period of only a few hours[19]. Though a direct comparison of these systems has yet to be done, the variety of peptide components present in the crude extract and the relatively high yields suggest the system is better optimized for bioproduction.

The second major section of this work, encompassing chapters 2 and 3, will focus on expanding recent efforts towards developing cell-free systems capable of generating metabolic products. Chapter 2 focuses on the investigation and engineering of the crude *E. coli* cell lysates through growth condition-based methods. This showed the use of shotgun proteomics and traditional analytical techniques as enabling technologies for rapidly engineering cell-free systems with an emphasis on the central metabolic precursor pyruvate.

Chapter 3 describes a Multiplex Automated Genomic Engineering (MAGE)-enabled protocol to reduce CFME extract contents through the incorporation of 6xHis tags on peptides detrimental to pyruvate pooling. This technique generates modified proteomes in order to produce engineered metabolic states with minimal impact on the cell in vivo thus enabling the ability to rapidly manipulate central metabolism without creating toxic phenotypes as often occurs from in vivo metabolic engineering. More broadly, this work reports on novel engineering tools capable of producing metabolic states traditionally impossible in living organisms.

Organization of Dissertation

The first chapter was originally published as the article “Computationally guided discovery and experimental validation of Indole-3-acetic acid synthesis pathways*”* and was undertaken to facilitate the annotation of uncharacterized proteins as part of the Plant-Microbe Interfaces (PMI) Science Focus Area (SFA) at Oak Ridge National Laboratory (ORNL)[29]. The second chapter is adapted from the article “Elucidating the potential of crude cell extracts for producing pyruvate from glucose” wherein we used metabolic proteomic analysis to measure the effect growth conditions can have on the function of cell-free extracts[30]. The third chapter “Engineering unconventional metabolism in cell-free systems” currently available on *bioRxiv* and explores a method of reducing cell-free proteomes to engineer cell-free metabolism. The supplementary information for each article is included in its respective chapter’s Appendix. References for each chapter have been consolidated at the end of the document.

Chapter 2

COMPUTATIONALLY GUIDED DISCOVERY AND EXPERIMENTAL VALIDATION OF INDOLE-3-ACETIC ACID SYNTHESIS PATHWAYS

A version of this chapter was originally published as:
C. Garcia, D., Cheng, X., L. Land, M., F. Standaert, R., L. Morrell-Falvey, J., and J. Doktycz, M. (2019) Computationally Guided Discovery and Experimental Validation of Indole-3-acetic Acid Synthesis Pathways. *ACS Chem. Biol. 14*, 2867–2875.

Abstract

Elucidating the interaction networks associated with secondary metabolite production in microorganisms is an ongoing challenge made all the more daunting by the rate at which DNA sequencing technology reveals new genes and potential pathways. Developing the culturing methods, expression conditions, and genetic systems needed for validating pathways in newly discovered microorganisms is often not possible. Therefore, new tools and techniques are needed for defining complex metabolic pathways. Here we describe an in vitro computationally-assisted pathway description approach that employs bioinformatic searches of genome databases, protein structural modeling, and protein-ligand docking simulations to predict the gene products most likely to be involved in a particular secondary metabolite production pathway. This information is then used to direct in vitro reconstructions of the pathway and subsequent confirmation of pathway activity using crude enzyme preparations. As a test system, we elucidated the pathway for biosynthesis of indole-3-acetic acid (IAA) in the plant-associated microbe *Pantoea sp.* YR343. This organism is capable of metabolizing tryptophan into the plant phytohormone IAA. BLAST analyses identified a likely three-step pathway involving an aminotransferase, an indole pyruvate decarboxylase, and a dehydrogenase. However, multiple candidate enzymes were identified at each

step, resulting in a large number of potential pathway reconstructions (32 different enzyme combinations). Our approach shows the effectiveness of crude extracts to rapidly elucidate enzymes leading to functional pathways. Results are compared to affinity-purified enzymes for select combinations and found to yield similar relative activities. Further, in vitro testing of the pathway reconstructions revealed the “underground” nature of IAA metabolism in *Pantoea* sp. YR343 and the various mechanisms used to produce IAA. Importantly, our experiments illustrate the scalable integration of computational tools and cell-free enzymatic reactions to identify and validate metabolic pathways in a broadly applicable manner.

Introduction

Metabolism is a complex network of interconnected chemical pathways responsible for an organism’s subsistence. In addition to macromolecules, small molecules (metabolites) are produced by cells. Primary metabolites (such as sugars, amino acids, lipids, nucleotides, and cofactors) provide energy, serve as building blocks for macromolecules or otherwise support core cellular functions. These molecules are common to many organisms, and their metabolism (biosynthesis and subsequent transformations) is relatively well understood. Secondary metabolites are small molecules made for other purposes, such as signaling and defense. They are highly diverse and produced by a limited number of organisms. While our understanding of primary metabolic networks has come a long way, the pathways by which the vast majority of secondary metabolites are created or utilized are poorly defined. For an overwhelming number of

microorganisms, there is little in the way of analytical evidence that a given pathway is either present or active[31,32]. This problem has expanded with the increase in genome-scale sequencing efforts. Genetic information on millions of proteins provide hints as to the existence of metabolic pathways, but robust methods for deciphering and confirming the metabolic potential hidden in this information are lacking[33]. Large collections of genomic data, such as the National Center for Biotechnology Information (NCBI) database, house over 40M protein sequences, the majority of which remain uncharacterized.[21] Automated annotations, while informative, are often incorrect and impede facile determination of metabolic capabilities [34,35]. The inability to confidently predict the function of an enzyme and its place in a metabolic pathway has become a consistent problem in the study and application of metabolic processes. Therefore, new approaches are needed for quickly and accurately elaborating the metabolic capabilities of microorganisms.

Current methods for defining metabolic networks can be slow and labor intensive. Given that methods for cultivation, pathway expression, and genetic manipulation are not always possible, conventional experimental approaches to the study of metabolic processes in newly identified microorganisms can be a challenging endeavor[36–39]. Previous work has sought to remedy this problem by leveraging computational tools to reveal the presence and potential function of a protein[40]. The combined use of bioinformatic software and structural data has resulted in significant strides in predicting enzymatic functions from

uncharacterized proteins[41–43]. For example, the addition of complementary data, such as genomic context and ligand docking analysis, has furthered this pursuit by demonstrating the functionality of multi-input predictions to generate testable hypotheses[44]. Such efforts have continued with recent advancements such as integrative pathway mapping wherein the function of a candidate enzyme and its potential metabolic pathways are predicted by combining information such as ligand docking, chemoinformatic analysis, genomic context, and chemical screening in a single analysis[45]. While these efforts have advanced significantly in recent years, determining the efficacy of these computationally-based approaches for defining protein function and pathway connectivity still requires experimental verification, and determination of protein activity remains a major bottleneck for effective gene annotation and pathway analysis.

To facilitate the definition and confirmation of metabolic pathways, we sought to develop a simple stepwise method that initially culls a large subset of proteins related to a pathway, using in silico methods, and tests the remaining potential pathways through scalable, in vitro biochemical experiments. Our combined computational and empirical approach towards pathway description consists of three steps. First, bioinformatic analyses of the genome or genome database of interest are performed using query enzymes described in the literature in order to find homologous enzymes and identify potentially complete pathways. In addition to removing the overwhelming majority of the genome, this bioinformatic step has the benefit of providing a loosely culled set of enzymes with potential activity based

on homology. Second, ligand-docking simulations are performed with protein crystal structures or computationally modeled structures to further cull the listed enzymes to those most likely to interact with their predicted substrates/intermediates. Third, small-scale, heterologous expression and *in vitro reactions in the crude extract are performed to examine the contribution of* individual enzymes to the predicted pathways, thus verifying pathway activity without the need for lengthy purification efforts. The resulting new understanding of the pathway and its component enzymes can then be used to refine gene annotation and make high-quality predictions for the presence of the same pathway in other organisms by clustering potential enzymes with the verified subset.

To test the effectiveness of using computationally guided discovery coupled with experimental validation in crude extracts, we examined the production of the phytohormone indole-3-acetic acid (IAA) from tryptophan in *Pantoea sp.* YR343 (YR343), a root colonizing member of the *Populus deltoides* plant-root microbiome[46]. This plant regulatory metabolite can be the product of a complex set of interconnected metabolic reactions. Notably, *Pantoea sp.* YR343 was shown previously to produce IAA in the presence of tryptophan, but the expected enzymes do not exist in a common operon [47]. This lack of common genomic context coupled with the annotation of many of the potential candidate enzymes confounds pathway determination. Further, as with many products of secondary metabolism, the number of potential pathways, as well as their interconnected

nature, complicates understanding of flux through the pathway and its potential control mechanisms. As described below, the combined computational culling and in-vitro verification approach effectively defines the functional capabilities of the component enzymes and the tryptophan to IAA pathway in *Pantoea* sp. YR343.

Results and Discussion

Bioinformatic Analyses Indicate Complex IAA Metabolism in Pantoea sp. YR343

Plants experience complex interactions with the microbial communities immediately surrounding and within their root systems[48]. The microbes in these systems are composed of bacteria, fungi, and archaea capable of influencing plant behavior through chemical signals[49]. *Pantoea* sp. YR343, a previously sequenced member of the *P. deltoides* plant-root community, was found to produce the plant phytohormone indole-3-acetic acid from tryptophan through an uncharacterized pathway[46,50]. Because some intermediates are shared by these pathways, there are seven interconnected but discrete potential pathways for IAA production from tryptophan (**Figure 2.1**)[47,51]. Prospective IAA biosynthetic pathways in YR343 were identified following a BLAST search using 11 different enzymes associated with these pathways acquired from bacterial and plant genomes (**Appendix Table 2.1**). *Pantoea* sp. YR343 contains at least one enzyme homolog for each of the known IAA pathways, but the only complete route from tryptophan to IAA is the IPA pathway, in which tryptophan is first converted to indole-3-pyruvate, then

indole-3-acetaldehyde, and finally IAA by an aminotransferase (Am-Trf), an indole pyruvate decarboxylase (IPDC), and an indole acetaldehyde dehydrogenase (IALDh), respectively (**Figure 2.1, Appendix Figure 2.4**). An E-value of 1e-25 was used as a cutoff value as it removed all enzymes from the known non-IAA producer, *E. coli*. Two of the three nodes in *Pantoea* sp. YR343's potential IAA pathway had multiple possible enzymes, specifically one aminotransferase, two IPDCs, and 16 dehydrogenases. Therefore, 32 distinct enzyme combinations could potentially complete a pathway for tryptophan to IAA conversion.

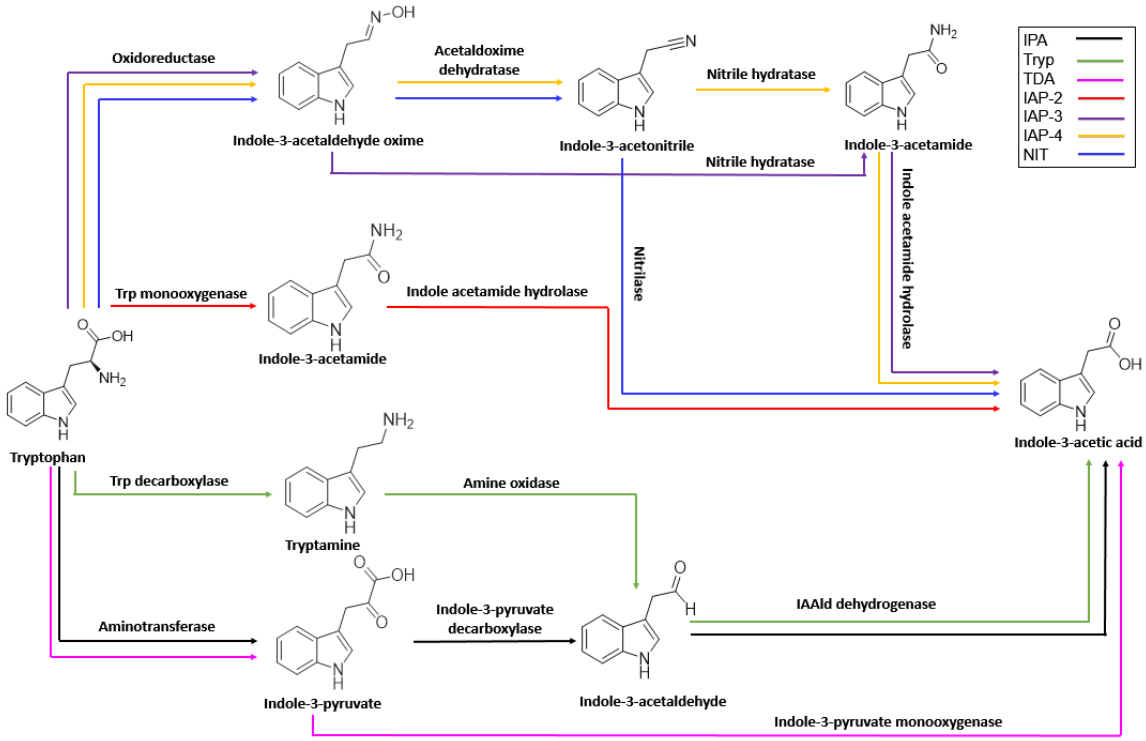

Figure 2.1. General metabolic model of the conversion of tryptophan to IAA. Enzymatic steps are written above the colored arrows and names of ligands and products underneath their molecular structure. The seven pathways found to produce IAA from tryptophan are color coded as follows: black, indole pyruvate pathway (IPA); green, tryptamine pathway (Tryp); magenta, tryptophan-dependent auxin biosynthesis pathway (TDA), red, 2-step Indole-3-acetamide pathway (IAP-2); purple, 3-step indole-3-acetamide pathway (IAP-3), yellow, four-step indole-3-acetamide pathway (IAP-4); blue, indole-acetaldoxime/indole-3-acetonitrile pathway (NIT).

Computational screening of ligand-docking interactions

To cull the list of potential gene products and resulting enzyme combinations possibly responsible for IAA production in *Pantoea* sp. YR343, ligand docking was used as a means of testing metabolic interactions and removing enzymes that are unlikely to bind their respective substrates/products. Given the unavailability of a crystal structure for any of the 19 enzymes, Phyre2 was used to generate

homology models and predict ligand binding sites **(Figure 2.2A)**[52]. Phyre2's top-scoring model and its predicted ligand-binding site were used as the basis for analyzing how the enzyme interacts with its potential substrate/product **(Figure 2.2B,C)**.

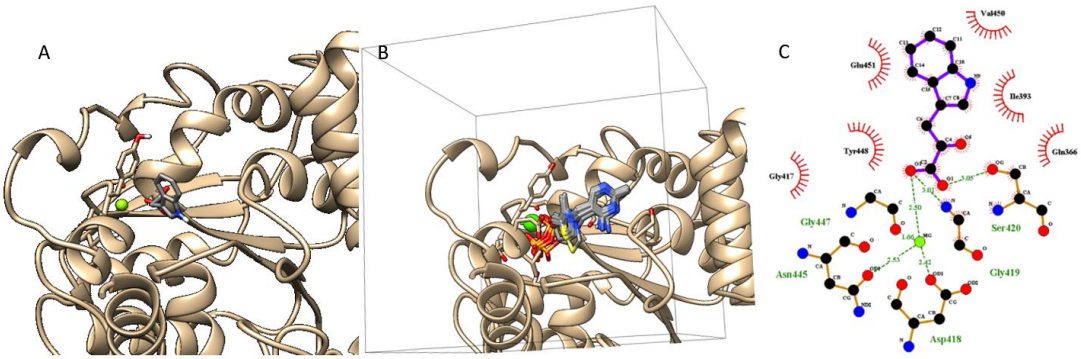

Figure 2.2. Homology modeling and ligand docking to pmi39_00059 IPDC. A. Docked indole-3-pyruvate in IPDC PMI39_00059. B.Docking search space (black box) and region predicted to be binding pocket by Phyre2. C. Schematic representations of docked ligands. Residues involved in hydrogen bonding interactions are shown as green dotted lines with the corresponding donor-acceptor distance is shown as a ball and stick model. Residues involved in van der Waals interactions with the ligand are shown with spikes.

The binding of a ligand in the binding pocket was visually inspected in n ≥ 5 independent simulations using Vina. A simple binary designation of binding or non-binding was then used to cull enzymes from being part of the IAA conversion pathway (**Appendix Figure 2.5**). Phyre2's 3DLigandSite was used to determine the position of the binding pocket based on 3DLigandSite's structural library of homologous peptides with bound ligands (**Appendix Figure 2.6**)[53]. Proteins capable of binding the expected ligands were kept in the pool of potential enzymes while those unable to bind were removed **(Appendix Table 2.2)**. Consequently,

as the aminotransferase successfully accommodated tryptophan in its binding pocket, it was retained. Only one of the two IPDCs was capable of binding indole-3-pyruvate, and 9 of the 16 potential IALDh enzymes were predicted to bind indole-3-acetaldehyde. Removing those enzymes that failed the docking step lowered the possible pathways to nine enzyme combinations.

Pathway verification

Our first set of in vitro experiments tested each of the nine predicted IAA enzyme combinations using heterologously expressed enzymes in clarified crude extracts (Am-Trf + IPDC + IALDh). Testing each combination for IAA production revealed that five of the potential enzyme sets produced IAA in relatively large amounts (>12 µg/mL) while two sets produced relatively low amounts (<3 µg/mL) and two other combinations showed no IAA production **(Figure 2.3A).** As expected, the control reactions containing just host extract derived from either BL21 Star(DE3) or BL21(DE3)pLysS did not produce IAA. Each node in the IAA pathway was subsequently tested by replacing it with a blank extract from BL21 Star(DE3) (**Figure 2.3B**). Control reactions without a heterologously expressed Am-Trf were performed using the most active IALDh enzymes as their production had previously been verified. In all cases, each pathway showed IAA production in the absence of Am-Trf. The IAA levels produced in these control reactions are relatively high and are likely due to the presence of aminotransferase activity in the *E. coli* extract. A similar experiment was performed but without the presence of the heterologously produced IPDC. In all cases the absence of IPDC prevented IAA

production, showing that IPDC is essential for IAA production in the bacterial extract. Similarly, the presence of a heterologously produced IALDh was essential for IAA production as in its absence no detectable IAA was produced. Of the nine IALDhs that passed the ligand docking test, five were able to complete the IAA pathway when combined with an Am-Trf and IPDC.

It was important to determine if the in-silico culling had eliminated viable enzymes. Accordingly, the remaining IPDC and seven of the IALDhs removed from the pool of enzymes during the ligand-docking analysis were similarly cloned, expressed, and tested for a potential role in IAA production. When combined with verified IALDhs, the culled IPDC, IPDC2, showed low IAA production, while two of the seven potential IALDhs were found to produce IAA at levels comparable to those of the non-culled enzymes **(Figure 2.3C)**. The IALDh reactions showed that ligand docking had a false negative rate of 28.6%, substantially lower than the false positive rate of 44.4%.

In order to determine the advantages and effectiveness of using crude extract preparations for pathway validation, enzymes with positive activity results in crude extracts were affinity purified using 6xHis tags. From the potential 10 enzymes, 5 were successfully affinity purified and used for IAA production experiments. Enzymatic combinations were prepared using AMTRF1, IPDC1, IPDC2, IALDH7, and IALDH16 **(Appendix Figure 2.7)**. The purified protein reactions showed the same activity trends as those that employed using crude extract combinations. Specifically, the absence of IPDC1 was shown to be necessary for IAA production

as its absence halted IAA production. IALDH16 maintained its relatively high-rate of activity as reactions containing IALDH16 produced 5.66 µg/mL IAA compared to the IALDH7's 2.04 µg/mL. The aminotransferase (AMTRF1) was able to carry out the expected transamination, though only ~1 µg/mL of IAA was produced **(Appendix Figure 2.8)**.

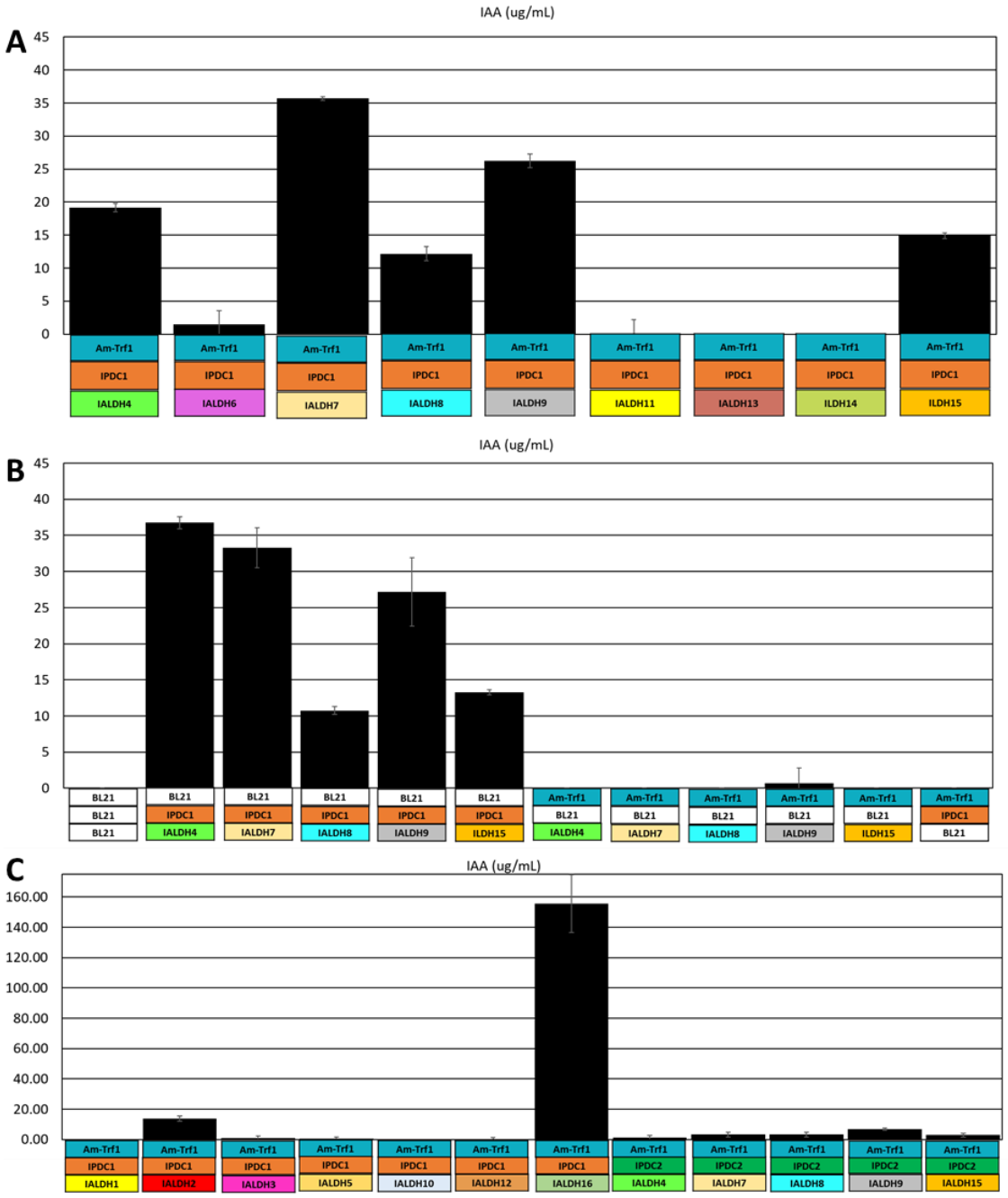

Figure 2.3. Modular assembly of potential IAA producing pathways. Complete pathways were obtained by using a single protein from each enzymatic step and measuring the terminal production of IAA. Gene loci are listed in Appendix Table 2.2. A. Only enzymes found to be bioinformatically related to a verified IAA enzyme and capable of docking their respective ligand was used to measure the initial concentration of IAA. Each reaction included a single aminotransferase, a single IPDC, and a single IALDh. B. Control reactions were performed to show the importance of individual metabolic steps. BL21 corresponds to an extract replaced with a BL21 Star DE3 extract containing no heterologously expressed protein. C. Culled enzymes were tested for activity to determine the rate of false negatives from Vina.

The plant microbiome consists of a complex network of chemical communications between a host and its microbial colonizers. IAA metabolism is highly relevant to plant-microbe interactions and serves as an important test case in developing a deeper understanding of both cellular metabolism and microbial communities. Even an apparently simple, three-step secondary metabolic pathway can lead to a large number of potential enzyme candidates and combinations that traditional methods cannot easily decipher. This work aims to refine an approach that can define relevant metabolic pathways, thus decreasing the resources necessary to establish secondary metabolic pathways.

An initial model for potential IAA metabolic pathways was created from the literature to generate BLAST hits. A relatively small set of enzymes involved in the known IAA pathways generated a large list of potential IAA pathway components in *Pantoea* sp. YR343. But, the comparison of potential pathway models and BLAST results provided a valuable culling step as it effectively removed enzymes from the querying pool and left the IPA pathway as the most likely generator of IAA. Further, BLAST analyses were key for IAA pathway description due to the lack of genomic context. Notably, of the nine enzymes that were eventually verified to be active in *Pantoea* sp. YR343's IAA metabolism, only three were found to exist within potential operons, all of which were unrelated to IAA metabolism (**Appendix Figure 2.5**).

In an effort to further cull the list of potential pathway components, homology modeling and ligand docking were employed. These emerging computational tools

can test a large set of potential candidates by virtue of the candidate's homology to the query set or ability to bind the predicted ligand. This study leverages previous efforts showing that the combined use of computational methods such as BLAST, homology modeling and ligand docking, while generating false positives and false negatives, is capable of identifying active proteins[54,55]. In this work, we used a readily accessible program, Vina, to dock ligands to the binding pockets of the protein homology models predicted by Phyre2 and 3DLigandSite. This process can be applied easily to a large number of candidate proteins and lends itself to further optimization and large-scale automation. As with many types of predictive analysis, binding affinity calculations are only estimates and may yield errors. Of the 19 proteins that passed the ligand docking test, seven were verified as having relevant substrate activity. Our analysis of each protein removed from the pool of enzymes during the ligand-docking step showed that of eight culled enzymes tested using Am-Trf1 and IPDC1, two of the culled IALDhs were found to produce IAA at levels comparable to those of the enzymes originally selected based on the ligand-docking **(Figure 2.3C)**. Overall, the presence of false negatives and positives is expected in this form of analysis and our results favor well when compared to similar studies[56]. Notably, less stringent conditions are capable of removing many false negatives by using higher energy confirmations, but this practice may exacerbate the acquisition of false positives[57,58]. While these accessible structural modeling and ligand docking tools can reduce search space

and simplify downstream in vitro experiments, further refinements are needed before confidently employing these tools on more complex pathways.

Future efforts to improve the performance of docking and homology modeling could lean more on an ensemble-based docking strategy akin to the Similarity Ensemble Approach (SEA), in which the binding ability is evaluated by a set of ligands (e.g., ligands similar to the target substrate/product and transition state structures), to obtain a more robust prediction of enzyme function[59]. Moreover, improvements can also be made by accounting for protein flexibility and by the use of methods such as molecular dynamics and quantum chemistry calculations that provide more accurate descriptions of the protein-ligand interactions[60,61]. At the moment, Vina's empirical scoring function relies on energetic factors to assign fitness. The use of automated computational tools will be essential for the evaluation of complex pathways and for keeping pace with genomic data.

A critical, complementary step in computationally-based pathway determination is experimental validation. Traditional methods employing genetic manipulation to verify activity are slow and not practical when analyzing more than one or a few enzymatic steps. Therefore, we employed a method of expressing and testing each of the predicted enzymes in a crude cell extract as a means to increase throughput. Traditional methods for enzyme characterization employ heterologous expression followed by affinity purification. Though affinity purification can lead to more definitive information, such as in terms of understanding reaction kinetics, it suffers due to poor success rates and low throughput[62–64]. Even high-throughput,

automated systems result in success rates as low as 20%, thus making affinity purification impractical for keeping up with expanding gene discovery[65,66]. In this work, reactions with purified enzymes were performed to evaluate the efficacy of using crude extracts. However, of the 9 different enzyme isolations attempted, only AMTRF1, IPDC1, IPDC2, IALDH7, and IALDH16 could be successfully purified. **(Appendix Figure 2.7)**. With these purified enzymes, activity trends were similar to those observed when using crude extracts of the proteins; the absence of IPDC1 still quenched IAA production and IALDH16 was still significantly more active than the other purified aldehyde dehydrogenase IALDH7. The time-consuming nature of purification is in contrast to the simplicity and effectiveness of using crude extracts. Here, the enzymes involved in IAA synthesis could be identified, expressed in soluble form and tested without the need for expansive troubleshooting. **(Appendix Figure 2.8)**.

Having defined the IAA pathway components employed by *Pantoea* sp. YR343, insights into the origin and function of the pathway can be gleaned. In the case of the initial aminotransferase step, the presence of an additional Am-Trf1 was not required for the reaction to progress. This is indicative of the well-known promiscuity of aminotransferases and suggests that the initial step in IAA metabolism is generally preserved among many organisms and used for other metabolic reactions[67]. Further testing of an affinity-purified version of the *Pantoea* sp. YR343 aminotransferase showed that the enzyme was capable of effectively catalyzing the reaction but only with high concentrations of the enzyme (**Appendix**

Figure 2.8). Additionally, we found that the IAA pathway in *Pantoea* sp. YR343 has a dependence on the IPDC PMI39_00059 (IPDC1). The importance of IPDC is evidenced by the high concentrations of IAA produced by each in vitro reaction containing IPDC1. This suggests that the predominant IAA pathway in YR343 utilizes PMI39_00059 and is a good target for further exploration of Pantoea sp. YR343 IAA metabolism. This result is substantiated by previous work wherein a full deletion of the PMI39_00059 gene generated an 80% drop in IAA[47].

The significant number of enzymes capable of catalyzing the dehydrogenation of indole-3-pyruvate to IAA emphasizes the need for rapid and effective in vitro tools. Sorting out the 16 different dehydrogenases by traditional genomic deletion approaches would be prohibitively difficult. While the lynchpin step in *Pantoea sp.* YR343's IAA metabolism is catalyzed by IPDC, the organism maintains several IALDh capable enzymes in its genome. Previous proteomic analyses of *Pantoea* sp. YR343 found that up to eight IALDhs could be detected at one time[47]. The redundancy of expression may indicate that *Pantoea* sp. YR343 maintains multiple pathways for IAA production. Interestingly, some of the most active enzymes found in this work were not found to be expressed in previous studies of *Pantoea* sp. YR343[47]. This may indicate that environmental conditions can control the expression of particular IALDh enzymes and consequently IAA. Regardless of other potential substrates for these IALDhs, these multiple reaction paths can be defined by the use of computational predictions when combined with validation using mixtures of crude enzyme extracts.

Definition of *Pantoea sp.* YR343's IAA pathway components uncovers its potential evolutionary development. The ability to find active enzymes, independent of genomic context, allows identification of those enzymes explicitly designed to carry out a function as well as those that exhibit flexibility in regard to substrate recognition. Underground metabolic functions describe potential side reactions of an enzyme and can serve as the basis for the emergence of new metabolic pathways without the need for significant evolutionary jumps[68]. In the case of *Pantoea* sp. YR343, the prevalence of specific enzymatic components related to IAA production and simultaneous lack of others shows the potential for underground metabolic reactions to generate novel metabolic functions. For *Pantoea* sp. YR343, the genes responsible for the three key conversion steps exist outside a common operon. IAA metabolism may, therefore, be an opportunistic phenotype generated through the crossover of a single gene in the form of IPDC or the mutation of a homologous IPDC gene. Using a crude extract approach to explore the metabolic potential of YR343, we were able to discern these promiscuous functions and provide an evolutionary context for the development of IAA metabolism in *Pantoea* sp. YR343.

This work has shown a rapid and scalable approach allowing for the identification and verification of active metabolic pathways in an organism by empirically generating functional annotations. The use of crude cell-extracts accurately predicted metabolic pathways despite the functional redundancy and lack of genomic context for the enzymes involved in IAA metabolism in the

organism *Pantoea* sp. YR343. We expect that improvements to computational tools will enhance the use of predictive analysis and the throughput of enzymatic discovery. As the breadth of genomic data continues to expand so too will the need to study such data without genetically tractable or even culturable organisms. The work presented demonstrates the effectiveness of crude extracts as a discovery and validation tool for both well-defined and underground metabolisms. Further, the development of high-throughput DNA synthesis and cell-free expression from eukaryotic cells could allow similar rapid explorations of metabolic pathways found in eukaryotic genomes and metagenomes. In-vitro enabled analysis such as that presented in this work, can help facilitate such studies by providing an accurate and rapid method of testing large search spaces in short amounts of time.

Methods

Pathway and gene identification

Following a literature search, a general model of IAA production from L-tryptophan was created from a set of query proteins in order to perform a BLASTP search against the *Pantoea sp.* YR343 translated genome (**Appendix Table 2.1, Figure 2.2**)[69]. At least one BLAST hit from the literature-derived query proteins was used to designate a protein as being related to IAA biosynthesis in *Pantoea sp.* YR343. Multiple queries were used when possible. As the goal of the BLAST analysis was to create a subset of proteins from the genome of interest for further

analyses, a single verified enzyme was deemed sufficient for the initial annotation. The BLAST analysis was performed by compiling a database of query proteins to BLAST against the *Pantoea sp.* YR343 genome. *Pantoea sp.* YR343's genome was downloaded from GenBank. The output was parsed by searching for potentially complete pathways in the genome using the general model created from the query sequences. The final set of pathways was obtained after setting the E-value cutoff at 1 × 10−25 in order to eliminate known non-IAA producers, in this case, *Escherichia coli*.

Ligand-docking simulations

Protein homology models were created using the Protein Homology/analogy Recognition Engine V 2.0 (Phyre2) by providing the target sequence from *Pantoea sp.* YR343 (**Appendix Table 2.2**).[70] The top-rated models were used, except in cases of low template identity in 3DLigandSite to predict the substrate-binding sites if this information was not immediately available from the template protein structures.[53] AutoDock Vina version 1.1.2 was employed to predict the interactions of each protein and its potential binding partners.[71] The search volume was set to 30×30×30 Å3 centered around the binding site predicted by 3DLigandSite; the exhaustiveness was set to 8, and maximum energy difference was set to 3 kcal mol^{-1} for each protein−ligand combination (**Figure 2.2A**). Successful binding partners were determined based on both the docking poses and docking score rankings (i.e., binding to the ligand- docking site predicted by 3DLigandSite); a set of binding and nonbinding representative examples can be seen in (**Appendix**

Figure 2.6). The lowest-energy pose was used to determine binding to the predicted site. Proteins predicted to bind their putative substrates were subsequently used for biochemical experiments and culled if no docked poses were identified by Vina.

Enriched extract preparation

One Shot TOP10 Chemically Competent E. coli (ThermoFisher) was used as the cloning strain for plasmid preparation. Each potentially IAA-related enzyme was amplified from *Pantoea sp.* YR343 and inserted into the NdeI/SalI site of pET-30a(+) growing on kanamycin (50 $\mu g mL^{-1}$)(**Appendix Table 2.3**). All primers were designed using NEBuilder (New England Biolabs) and purchased from IDT. Crude cell-free extracts were prepared by culturing E. coli BL21 Star (DE3) in 2×YPTG (16 g L^{-1} tryptone, 10 gL^{-1} yeast extract, 5 g L^{-1} NaCl, 7 g L^{-1} KH2PO4,3 gL^{-1} K2HPO4, 18 g L^{-1} glucose). Cultures of 50 mL volumes were grown using 250 mL baffled flasks at 37 °C shaking at 250 rpm. Induction was performed using 0.1 mM IPTG at OD600 = 0.6−0.8 and harvested after growing for 4 h at 30 °C. No antibiotics were used during growth. Each culture was harvested by centrifugation at 5000g for 10 min and washed twice with S30 buffer (2 g L^{-1} magnesium acetate, 14.05 g L^{-1} potassium glutamate, 0.154 g L^{-1} dithiothreitol (DTT), and 1.81 g L^{-1} Tris-acetate, pH 8.2). After the final wash, the cell pellets were weighed, flash-frozen in liquid nitrogen, and stored at −80 °C. Cell extracts were made by thawing and resuspending the pellet in 0.8 mL of S30 buffer per gram of wet cell weight before sonicating with 530 J mL^{-1} of suspension at 50% amplitude while in ice

water. After sonication, the lysed cells were centrifuged twice at 4°C for 10 min at 21,100g. The clarified lysate was flash-frozen, and stored at −80 °C. The BL21 Star (DE3) strain was used as the base expression strain. Plasmids that did not produce soluble protein in BL21 Star (DE3) were moved to BL21(DE3)/pLysS but were otherwise prepared in the same manner. Both pmi39_00977 and pmi39_04201 successfully expressed in BL21(DE3)/pLysS. After preparing each extract, 8 ng of the soluble crude extract was loaded onto an SDS-PAGE gel in order to verify the expression of each enzyme. His-tag purified enzyme production required varying growth conditions outlined in the Supporting Information (**Appendix Figure 2.7**).

In vitro reactions

In vitro IAA synthesis reactions were prepared by combining 50 mM L-tryptophan and ∼5mgmL−1 of each enriched extract in a 30 µL volume. Control reactions that omitted putative reaction steps had volume shortage made up using cell extract from E. coli. The reactions were then placed in a 28 °C incubator shaking at 250 rpm for 24 h. Following incubation, each reaction was placed on ice and deactivated by adjusting the pH to 2.0 with HCl. IAA was extracted by the addition of 500 µL of ethyl acetate and vigorously vortexed. Each sample was then incubated on ice for 5 min and briefly centrifuged to separate aqueous and organic layers. A 400 µL portion of the organic layer was removed and dried in an analytical vial with argon gas. Vials were stored at 4 °C before analysis. Stored samples were resuspended in 500 µL of water before injection. Reaction components and

conditions for purified enzyme reactions are described in detail in **Appendix Figure 2.8**.

Analysis of IAA production

Quantitative analysis was performed by injecting 100 µL of suspended reaction into an Agilent 1260 HPLC instrument equipped with an Agilent ZORBAX Eclipse Plus C18 column and a diode array detector set at 265 nm. The mobile phase comprised A, 0.08% trifluoroacetic acid in water; and B, acetonitrile. Analytes were eluted at a flow rate of 1 mL min^{-1} with the initial eluent composition of 5% B held for 3 min, followed by a step to 30% B, a 5 min linear gradient to 45% B followed by a 5 min hold, a step to 75% B followed by a 1 min hold, and a step to 5% B followed by a 6 min hold. A calibration curve was prepared using pure IAA. Representative spectra are shown in the Supporting Information (**Appendix Figure 2.9**). All chemicals were acquired from Sigma-Aldrich. Error bars were calculated using the standard deviation from n ≥ 3 independent reactions.

Appendix

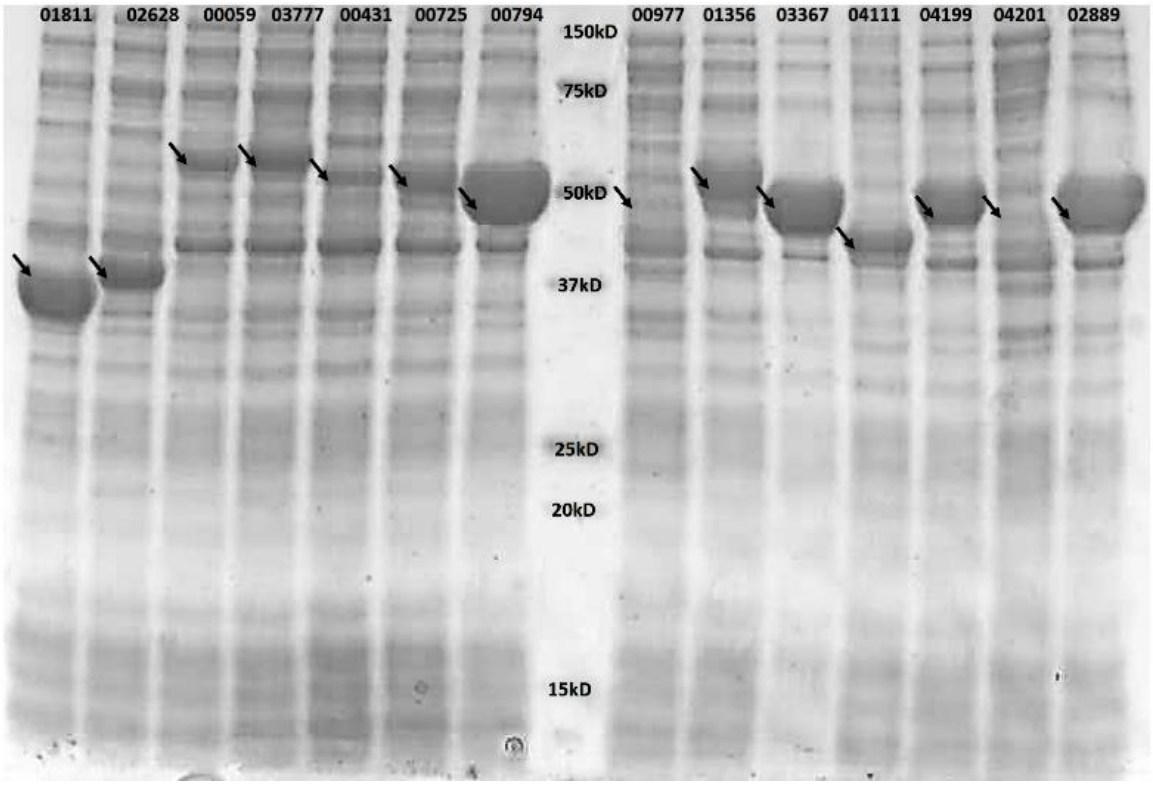

Figure 2.4. Potential IAA-related proteins were expressed in *E. coli* and tested for solubility. Arrows mark the presence of the resultant gel band.

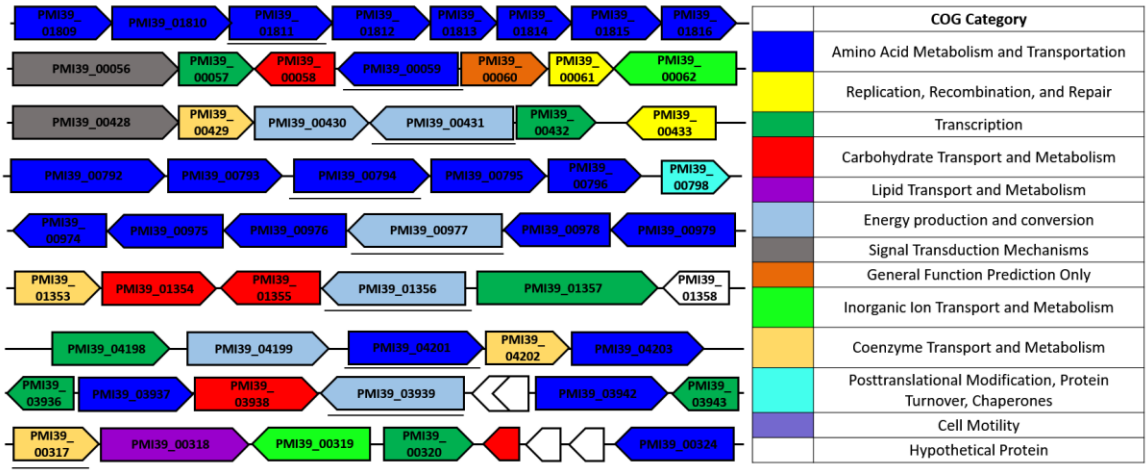

Figure 2.5. Gene neighborhoods for each of the YR343 enzymes expressed and found to be part of an IAA reaction. Enzymes within the genomic region found to produce IAA are underlined.

Table 2.1. Representatives for enzymes used in BLAST analysis. Each enzyme's name and source are shown, as well as the number of hits found in *Pantoea* sp. YR343's genome at a cutoff of 1e-25.

Enzyme	Query Source	YR343 Representatives
aminotransferase	>gi|329738252|gb|AEB97257.1| HisC1 [Azospirillum brasilense][72]	1
indole-3-pyruvate decarboxylase	>gi|508716|gb|AAC36886.1| indole-3-pyruvate decarboxylase [Azospirillum brasilense Sp245][73] >gi|837358496|gb|AKM88529.1| indolepyruvate decarboxylase [Enterobacter cloacae][74]	2
indole-3-acetaldehyde dehydrogenase	>tr|A0A060DWZ5|A0A060DWZ5_AZOBR Aldehyde dehydrogenase OS=Azospirillum brasilense GN=ABAZ39_26155 PE=3 SV=1[75] >gi|757149327|ref|WP_042703524.1| aldehyde dehydrogenase [Azospirillum sp. B506] >gi|1658175|gb|AAC49575.1| indole-3-acetaldehyde dehydrogenase [Ustilago maydis][51]	16
indole-3-pyruvate monooxygenase	>gi|75213680|sp|Q9SZY8.1|YUC1_ARATH RecName: Full=Probable indole-3-pyruvate monooxygenase YUCCA1; AltName: Full=Flavin-containing monooxygenase YUCCA1 >gi|910738333|dbj|BAS07733.1| probable indole-3-pyruvate monooxygenase YUCCA3 [Arthrobacter sp. Hiyo4]	0
Trp-monoxygenase	>AGL87350.1 tryptophan 2-monooxygenase IaaM [Pseudomonas protegens CHA0][76] >gi|4836569|gb|AAD30489.1|AF126446_4 tryptophan monooxygenase (plasmid) [Agrobacterium fabrum str. C58]	0
Indole acetamide hydrolase	>gi|151292|gb|AAA25853.1| indoleacetamide hydrolase [Pseudomonas syringae][76]	6
oxidoreductase	>gi|12644083|sp|O81346.2|C79B2_ARATH RecName: Full=Tryptophan N-monooxygenase 1; AltName: Full=Cytochrome P450 79B2; AltName: Full=Tryptophan N-hydroxylase 1[77] >gi|75319827|sp|Q501D8.1|C79B3_ARATH RecName: Full=Tryptophan N-monooxygenase 2; AltName: Full=Cytochrome P450 79B3; AltName: Full=Tryptophan N-hydroxylase 2[77]	0
nitrile hydratase	>gi|426268469|gb|AFY20546.1| nitrile hydratase alpha subunit [Pseudomonas sp. UW4][73] >gi|426268470|gb|AFY20547.1| nitrile hydratase beta subunit [Pseudomonas sp. UW4][73]	0
acetaldoxime dehydratase	>gi|6855269|dbj|BAA90461.1| phenylacetaldoxime dehydratase [Bacillus sp. OxB-1][78] >gi|46917226|dbj|BAD17969.1| aldoxime dehydratase [Rhodococcus erythropolis][79]	0
Trp-decarboxylase	>EDU35915.1 pyridoxal-dependent decarboxylase domain protein [Clostridium sporogenes ATCC 15579] >4OBV_A Chain A, Ruminococcus Gnavus Tryptophan Decarboxylase Rumgna_01526 (alpha-fmt)[80]	1
amine oxidase	>Azospirillum brasilense Az39: ABAZ39_10795 Amine Oxidase >Azospirillum brasilense Sp245: AZOBR_50011 Amine oxidase	0

Table 2.2. List of template structures used for protein homology modeling. Target sequence locus tag designates the sequence address provided by the Joint Genome Institute's Integrated Microbial Genome and Microbiomes (JGI IMG/M) system. Templates used by Phyre2 are available in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB). Molecular models for ligands were downloaded from the ZINC database.

Target Sequence/Locus	Template Structure	Template Sequence Identity	Target Protein	Ligand
PMI39_01811	1FG7	74	aminotransferase	L-tryptophan
PMI39_00059	1OVM	57	indole-3-pyruvate decarboxylase	indole-3-pyruvate
PMI39_03777	1YI1	40	indole-3-pyruvate decarboxylase	indole-3-pyruvate
PMI39_00313	2VE5	74	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_00317	2VE5	79	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_00354	2W8Q	51	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_00431	4O5H	37	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_00617	2W8Q	52	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_00725	4GO4	42	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_00794	5U0M	62	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_00977	5U0M	60	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_01356	4O5H	41	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_02889	2VE5	76	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_03367	4ZZ7	57	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_03939	4F9I	34	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_04111	4F9I	28	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_04199	3EK1	62	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_04201	1BXS	37	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde
PMI39_04236	3VZ0	49	indole-3-acetaldehyde dehydrogenase	indole-3-acetaldehyde

Table 2.3. Enzymes identified through a BLAST search of *Pantoea* sp. YR343's genome. A cutoff of 1e-25 was used for BLAST analysis. Locus designates the relative position of each enzyme in the YR343 genome. Enzyme title indicates the designation as an aminotransferase (Am-Trf) indole-3-pyruvate decarboxylase (IPDC), or indole-3-acetaldehyde dehydrogenase (IALDh). Ligand docking indicates the expected ability of the ligand to bind to the protein, (+) if successfully docked, (-) if unsuccessfully docked. IAA production designates a protein as being part of an active IAA pathway.

Enzyme Title	Locus	Amino Acid Length	E-Value	Ligand Docking	IAA Production
Am-Trf1	PMI39_01811	358	1.0E-32	+	+
IPDC1	PMI39_00059	550	2.0E-38	+	+
IPDC2	PMI39_03777	574	6.0E-26	-	+
IALDH1	PMI39_00313	146	2E-26	-	-
IALDH2	PMI39_00317	326	3E-74	-	+
IALDH3	PMI39_00354	489	5E-91	-	-
IALDH4	PMI39_00431	472	2E-77	+	+
IALDH5	PMI39_00617	480	6E-87	-	-
IALDH6	PMI39_00725	506	4E-100	+	-
IALDH7	PMI39_00794	488	2E-42	+	+
IALDH8	PMI39_00977	492	1E-46	+	+
IALDH9	PMI39_01356	506	3.0E-104	+	+
IALDH10	PMI39_02889	490	1.0E-111	-	-
IALDH11	PMI39_03367	501	1E-67	+	-
IALDH12	PMI39_03939	478	6E-91	-	-
IALDH13	PMI39_04111	458	6E-56	+	-
IALDH14	PMI39_04199	485	2E-88	+	-
IALDH15	PMI39_04201	489	1E-93	+	+
IALDH16	PMI39_04236	457	3E-53	-	+

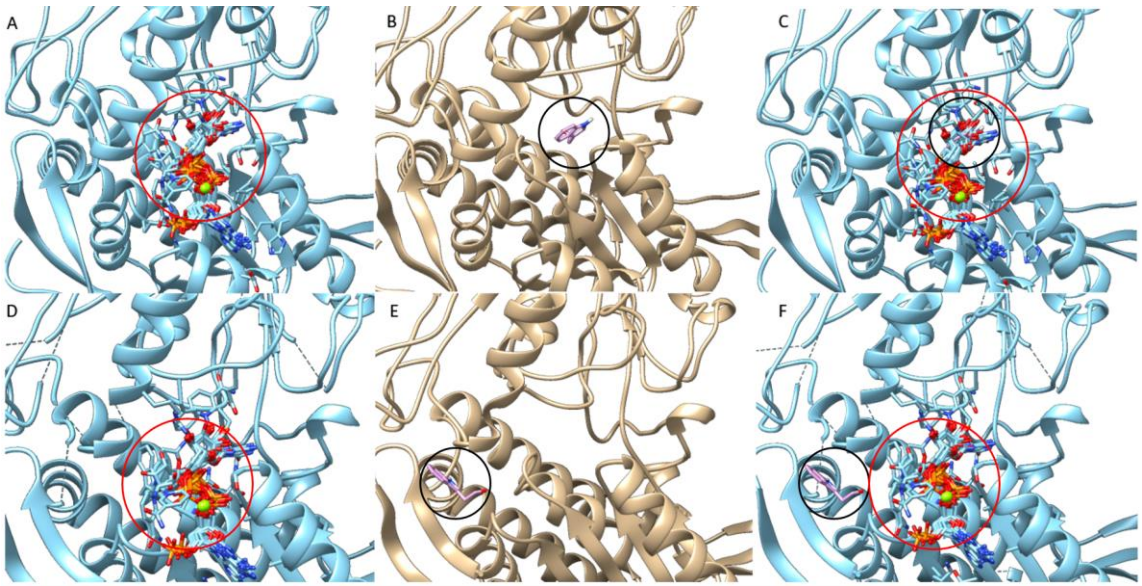

Figure 2.6. Representative examples of binary designation used to describe binding and non-binding ligands. Designation based on lowest energy pose found by AutoDock Vina. Binding pockets are outlined in a red circle and ligands are outlined in a black circle for highlighting purposes. (A) 3DLigandSite prediction of ligand binding pocket for PMI39_01356. (B) Lowest energy conformation for binding indole-3-pyruvate (purple) to PMI_01356 found by AutoDock Vina. (C) Overlap of lowest energy binding pose for binding indole-3-pyruvate (purple) and 3DligandSite for PMI39_01356. It should be noted that the ligand is not visible as a result of the overlap with molecules in the binding site. D. 3DLigandSite prediction of ligand binding pocket for PMI39_04236. E. Lowest energy conformation of non-binding indole-3-pyruvate (purple) to PMI39_04236 found by AutoDock Vina. F. No overlap is seen between the lowest energy binding pose of indole-3-pyruvate (purple) and 3DligandSite for PMI39_04236.

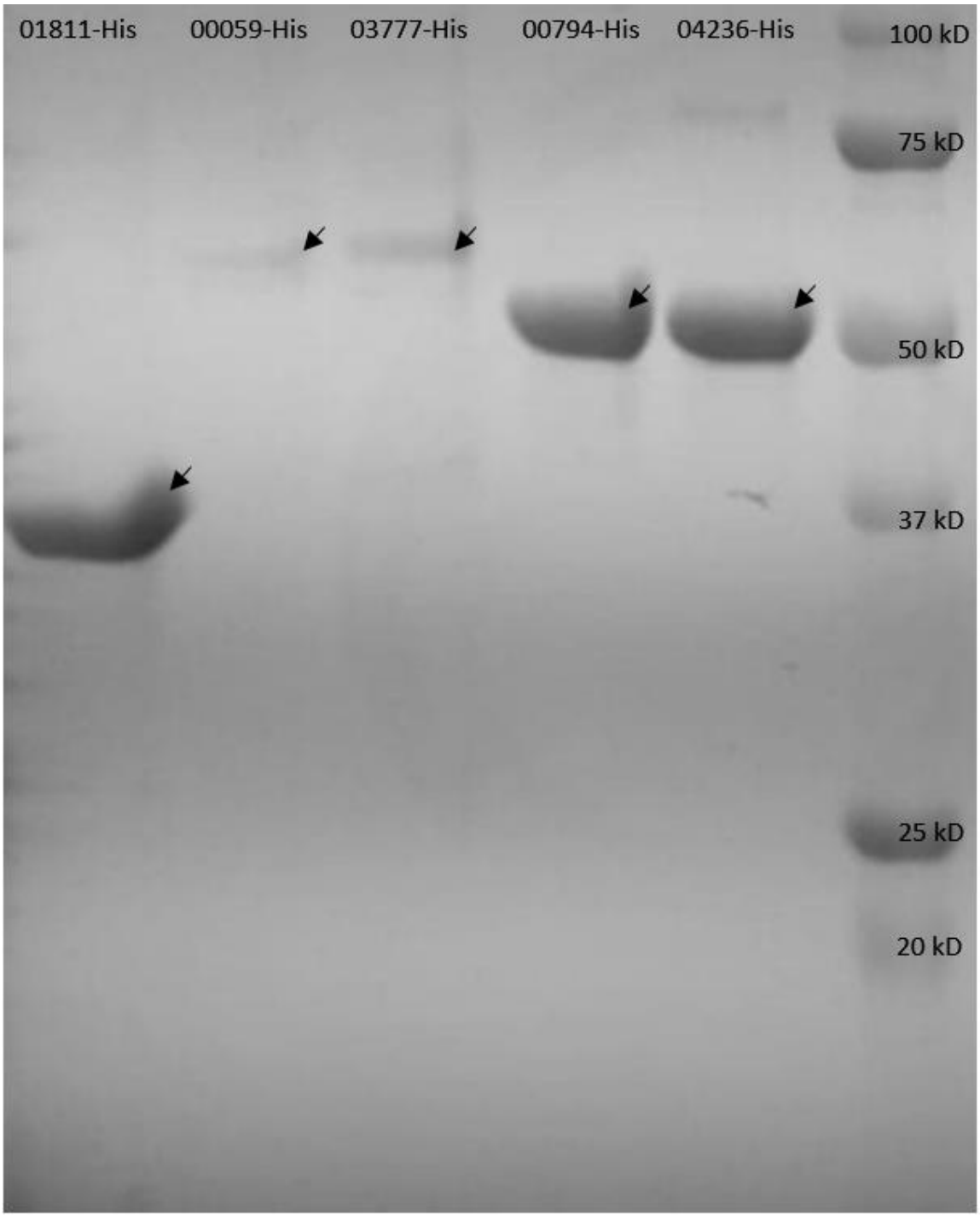

Figure 2.7. SDS-PAGE gel of His-Tagged IAA-related proteins expressed in *E. coli* and purified. Arrows mark the presence of the resultant gel band.

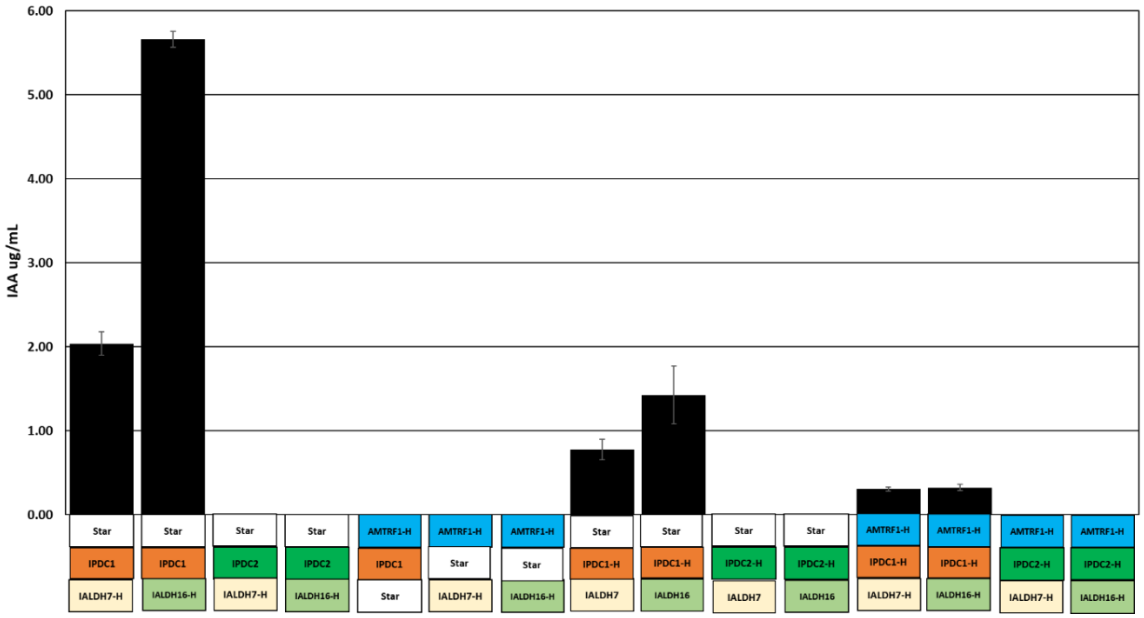

Figure 2.8. Assembly of purified IAA producing enzymes. Purified enzymes were also combined with crude extracts including a single aminotransferase, a single IPDC, and a single IALDh in a crude or 6x-His purified form. Purified versions of each protein are indicated by the addition of an H to enzyme name. “Star” surrounded by a white box corresponds to a BL21 Star DE3 extract containing no heterologously expressed protein. “Star” extracts, were loaded to a final concentration of 5.94 mg/mL; AMTRF1 was loaded to a final concentration of 3.8 mg/mL; IPDC1 extracts were loaded to a final concentration of 1.054 mg/mL; IPDC2 was loaded to a final concentration 1.054 mg/mL; IALDH7 and IALDH16 were loaded to a final concentration of 4.92 mg/mL; AMTRF1-H, IPDC1-H, IPDC2-H, IALDH7-H, IALDH16-H, were all loaded to a final concentration of 0.9 mg/mL. All reactions had final L-tryptophan concentrations of 50 mM and were incubated at 28° C for 24 hours. Each reaction was carried out in 50 µL volumes except those comprised of entirely purified enzymes which were 166 µL.

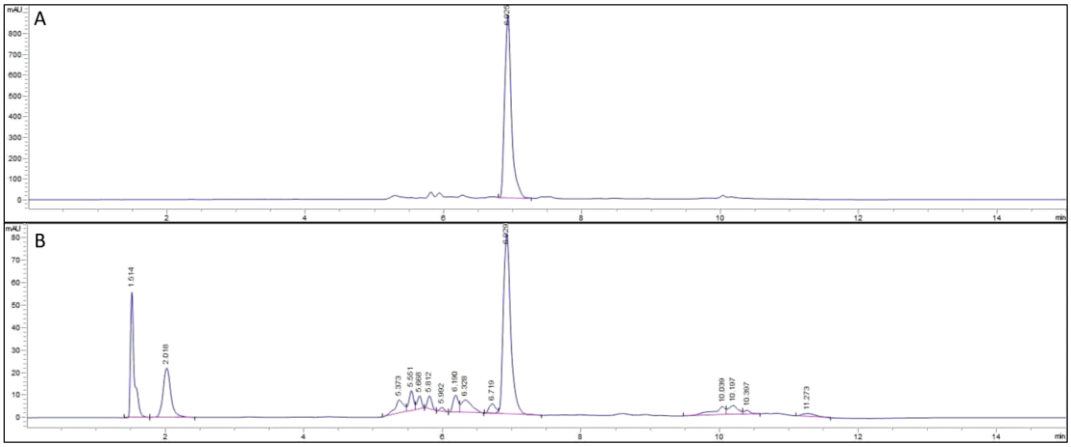

Figure 2.9. HPLC analysis of IAA. A. Injection of 5 µg/mL purified IAA. B. Representative cell-free reaction following injection into HPLC.

Chapter 3

ELUCIDATING THE POTENTIAL OF CRUDE CELL EXTRACTS FOR PRODUCING PYRUVATE FROM GLUCOSE

A version of this chapter was originally published as:
Garcia, D. C., Mohr, B. P., Dovgan, J. T., Hurst, G. B., Standaert, R. F., and Doktycz, M. J. (2018) Elucidating the potential of crude cell extracts for producing pyruvate from glucose. *Synth. Biol. 3*, 1–9.
DCG, BPM, 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. DCG and BPM wrote the manuscript with assistance from RFS and MJD.

Abstract

Living systems possess a rich biochemistry that can be harnessed through metabolic engineering to produce valuable therapeutics, fuels and fine chemicals. Despite 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.

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. 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[81,82]. Coupled with systems biology tools for flux balance analysis and elementary mode analysis, in vitro systems present a potent platform for bioproduction[83]. The absence of a cell wall and membrane facilitates the exchange of substrate to, and product from, the system and simplifies reaction workup.

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[84].

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[12,85]. While recent efforts in co-purification of full reaction cascades have reduced costs, any process utilizing bulk purified proteins remains expensive[86,87]. 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[24]. Additionally, the 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[6,20]. 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[88,89]. 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[8]. Proteomic analyses coupled with metabolite measurements can be used to identify and characterized the biochemical pathways capable of being supported by crude extracts.

Results and Discussion

Given that the metabolic capabilities of a cell-free extract result from the active proteome, we hypothesized that changes to growth conditions prior to the preparation of cell extracts would create shifts in the protein content and the resulting metabolic abilities of the crude extracts. To investigate 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[90]. 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 OD_{600} of 8-10. Cells grown in this media were collected at both early, (OD_{600} 2.8) and mid-log phase (OD_{600} 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[91]. 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[88,89]. 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[94].

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[95]. 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[96]. 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 3.10**). 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.

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 productionn. 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 3.4. Figure 3.11 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[97]. 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 3.11**). While the DF extract condition contains the fructose/mannitol-specific IIBC protein, the YT-E extract proteome uniquely contains the glucose-specific BCII 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[98]. GSEA was performed as pairs of comparisons but some enrichments were shared amongst different comparisons and were consolidated (**Table 3.4**).

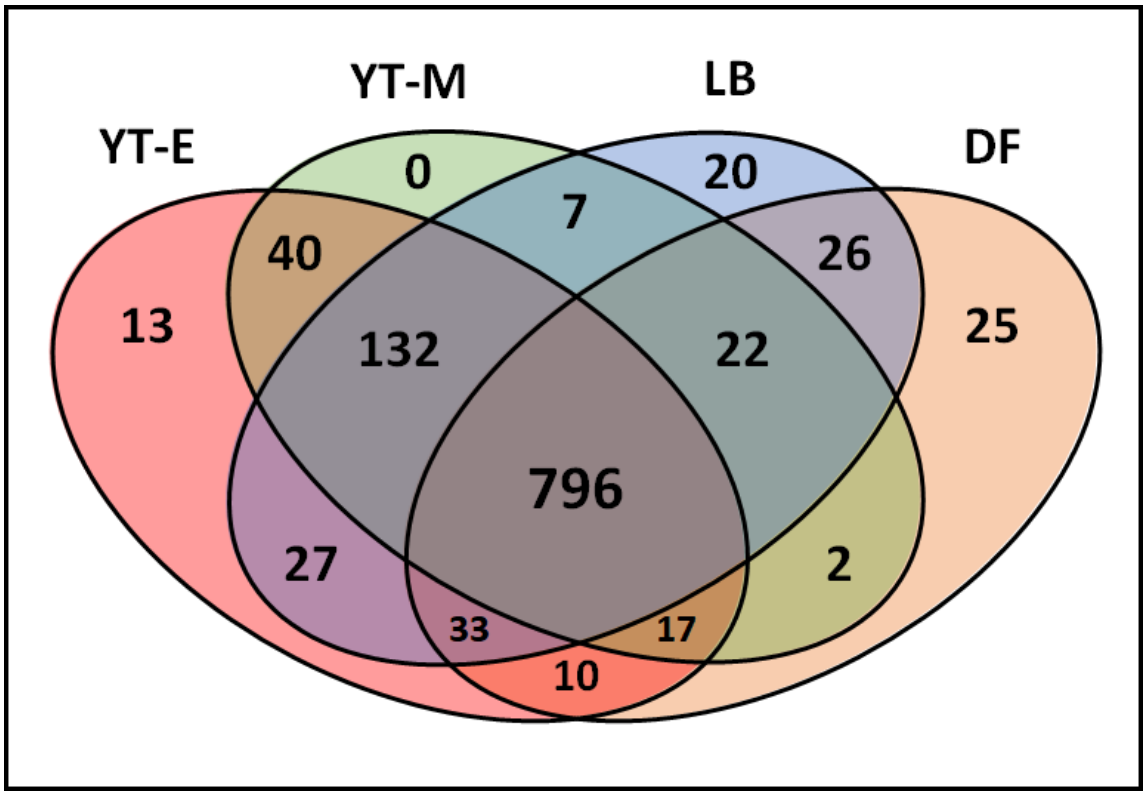

Figure 3.10. Venn diagram of differences in protein content from extracts produced from 4 different Growth conditions. Differences in protein content reported after filtering out low abundance proteins. There is a core set of 796 proteins present in all four conditions. The different conditions contain 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.

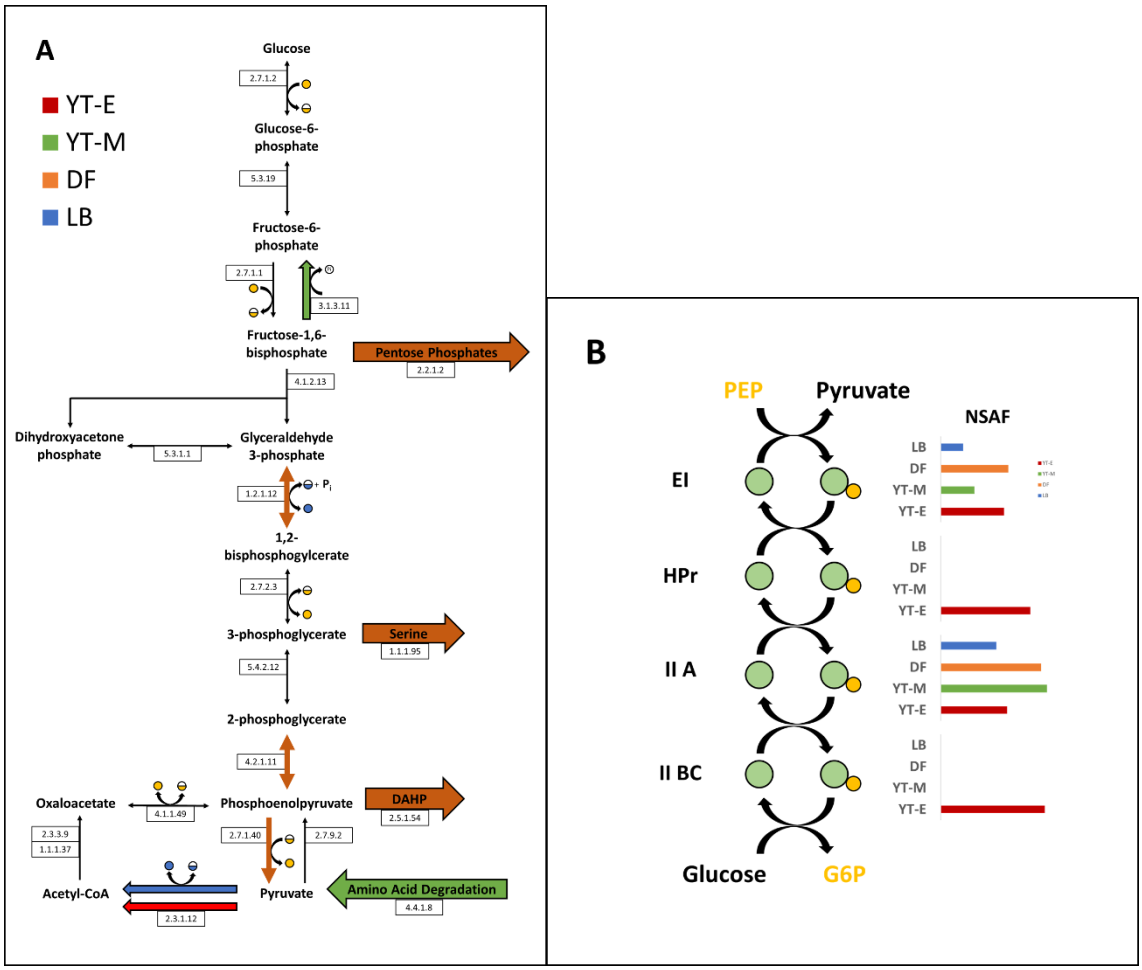

Figure 3.11. Proteomic assessment of the potential in vitro metabolic fates of glycolytic intermediates**.** A. Enzymatic steps are represented by Enzyme Commission (EC) numbers. Each EC number and reaction arrow corresponds to at least one protein detected in all growth conditions. Upregulated steps 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 the YT-E condition. Enrichment is represented by NSAFs.

Table 3.4. Summary of gene set enrichment analysis. Results are 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	structural constituent of 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	DNA binding transcription factor activity	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

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 h^{-1} over a 24 hour period (**Figure 3.12**).

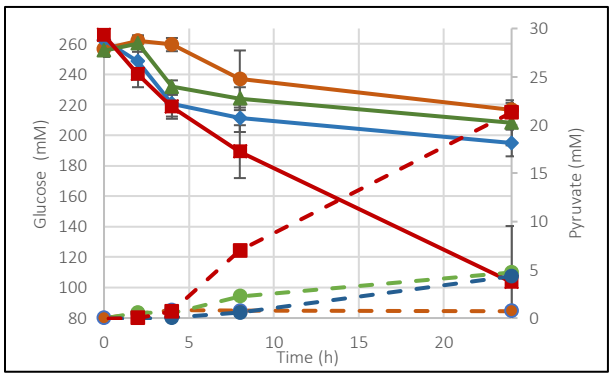

Figure 3.12. Simultaneous measurements of glucose and pyruvate at various time points over a 24-hour period. Extracts are shown here as orange, blue, green and red colored symbols and lines for DF, LB, YT-M, and YT-E growth conditions, respectively. Data and standard deviation for the time course reactions were acquired using n=3 biological replicates. Solid lines indicate glucose time courses and dashed lines indicate those for pyruvate.

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.

Table 3.5. Percent glucose consumed and the percent glucose converted to pooled pyruvate. Conversion amounts were determined using n=3 biological replicates. Data acquired 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. [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.

Extract	Percentage of Glu Consumed	Percentage of Consumed Glucose Converted to Pooled Pyruvate[a]	Percentage 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%

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 3.13)**. As suggested from the proteomic analyses, sink pathways for 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 3.5**). 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, 0.96 mM $h^{-1,}$ and 0.93 mM h^{-1} respectively, despite the YT-E extract producing a larger pool of pyruvate after 24-hours.

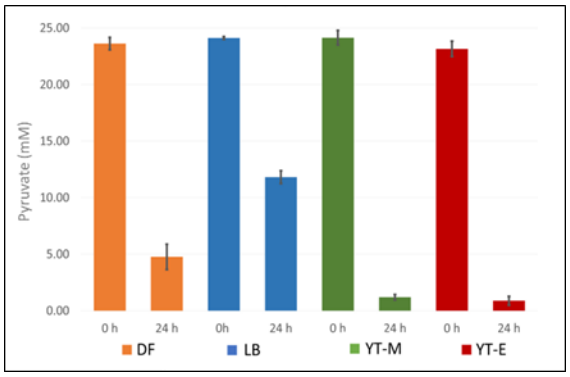

Figure 3.13. Simultaneously measurements of glucose and pyruvate at various time points over a 24-hour period. Extracts are shown here as orange, blue, green, and red colored symbols and lines for DF, LB, YT-M and YT-E growth conditions, respectively. Data and standard deviation for the time course reactions were acquired using n=3 biological replicates. Solid lines indicate glucose time courses and dashed lines indicate those for pyruvate.

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.

The metabolic potential of crude extract preparations and their use for metabolite production can be assessed through an 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[90]. 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 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 can 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[88]. 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 3.11, Supplemental Table S3**). 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 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 0.82 mM h^{-1}, 0.71 mM h^{-1}, 1.14 mM h^{-1}, and 1.82 mM h^{-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[99]. 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[100]. 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[101]. 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 the 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[92]. 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[102]. Crude extracts made from high-yielding in vivo pyruvate production strains represent another opportunity to use in vitro synthetic biology to enable metabolic engineering[103,104]. 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.

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 the production of useful metabolites.

Methods

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 gL^{-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 OD_{600} 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 OD_{600} 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.

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.

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^{-1} at 35 °C for 25 min.

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 the 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 × 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 (Washburn et al., 2001). 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 (Ma et al., 2009; Tabb et al., 2007). KEGG Orthologies and Enzyme Commission numbers were assigned by BlastKOALA (Kanehisa et al., 2016). Full tables of detected proteins, tryptic peptides and KEGG orthology

assignments are deposited in **Supplemental Tables S1** and **S2**. Descriptions of these tables and their legends are supplied in the appendix.

Statistical analysis

Three biological replicates were used for all HPLC measurements. Error bars in figures represent ± 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[8]. Gene set enrichment analysis (GSEA) was performed using the tools designed by Subramanian *et al.* (Subramanian et al., 2005). In brief, GSEA was performed using gene ontology (GO) terms and Uniprot pathway and super pathway 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.

Appendix

The following contents can be found in the original publication of this work as, Legend for Supplemental Table S1, Legend for Supplemental Table S2, Legend for Supplemental Table S3 and relate to the supplementary data therein.

Supplemental Table S1 (prot_supp_table.xlsx) includes a complete listing of the 1,763 proteins (one per row) identified by full proteome characterizations of crude cell extracts prepared from four different growth conditions. Columns in this table are:

Protein: protein identifier from FASTA file used for Myrimatch searches (see Methods: 2.4 Proteomics)

A set of 7 columns listing quantities measured or calculated from each LC-MS/MS measurement; “NA” indicates that a protein was not identified in that LC-MS/MS analysis:

- **SqC**: Sequence Count, the number of identified tryptic peptides for the protein, with different charge states of the same peptide counting as 1 sequence. Each identified modified form of a peptide is counted as a separate sequence.
- **uSqC**: number of identified sequences that are unique within the FASTA file to this protein
- **SpC**: Spectrum Count, the number of identified tandem mass spectra for this protein
- **uSpC**: Spectrum Count of peptides that are unique within the FASTA file to this protein
- **dSpC**: Distributed Spectrum Count, modified to account for peptides shared with other proteins[105].
- **PercentCoverage**: fraction of protein sequence covered by detected tryptic peptides
- **NSAF**: Normalized Spectral Abundance Factor: corrected and normalized estimate of protein abundance. NSAF sums to 1 across all proteins in a single run[106].

Headers for each of these 7 columns include the sample identifier:

- YT-E 1, YT-E 2, YT-E 3: LC-MS/MS biological replicates of standard *Escherichia coli* cell-free protein synthesis crude extract. Extract was grown on 2xYTPG media and collected during early-log phase.
- YT-M 1, YT-M 2, YT-M 3: LC-MS/MS biological replicates of *E. coli* crude extract. Extract was grown on 2xYTPG media and collected during mid-log phase.
- DF 1, DF 2, DF 3: LC-MS/MS biological replicates of *E. coli* crude extract. Extract was grown on M9 minimal media with fructose and collected during mid-log phase.
- LB 1, LB 2, LB 3: LC-MS/MS biological replicates of *E. coli* crude extract. Extract was grown on M9 minimal media with fructose and collected during mid-log phase.

Other_prots: Proteins in the FASTA file that share identified tryptic peptide sequence(s) with this protein Length: number of amino acids in the protein

Description: protein description from identifier line of the protein FASTA file

KO: K number orthology term assigned by BLASTKOALA[96]

Score: weighted sum BLAST bit scores for each K number group

Supplemental Table S2 (pep_supp.xlsx) lists tryptic peptides identified in full proteome characterizations of crude cell extracts prepared from four different growth conditions. Columns in this table are:

Sequence: amino acid sequence of the peptide. A number in square brackets indicates a modification in mass to the preceding amino acid.

Z: charge state of the peptide.

A set of 3 columns listing quantities measured or calculated from each LC-MS/MS measurement; “NA” indicates that a peptide was not identified in that LC-MS/MS analysis:

- **Q**: maximum Q value (IDPicker false discovery estimate for peptide-spectrum matches) among identifications of this peptide (peptide-spectrum

matches) that pass filters documented in Methods: Proteomics data analysis.

- **Precursor m/z**: mass-to-charge ratio of the precursor ion that was fragmented, averaged across the spectra identified for this peptide**.**
- **Spectra**: Number of tandem mass spectra matched with this peptide sequence by Myrimatch.

Headers for each of these 3 columns include a sample identifier, listed above for Supplemental Table 1.
Proteins: Proteins that contain this amino acid sequence

Supplemental Table S3 (glycolytic_sink_statistics.xlsx) lists each enzyme in *E. coli* proteome UP000002032, as both an EC number and proteome accession number, that consumes or produces an intermediate of glycolysis. Columns in this table are:

Glycolytic intermediate and Enzyme Commission number: lists the intermediate produced or consume and the EC number of the enzyme involved.

Protein accession number: the accession number for the protein in the provided proteome.

Average NSAF: as in **Table S1**. NSAFs were omitted for proteins that fell beneath the abundance threshold or were not detected.

ANOVA: Benjamini-Hochberg corrected p-values for One-way ANOVA. Values were omitted for corrected p-values > 0.05.

Chapter 4

A LYSATE PROTEOME ENGINEERING STRATEGY FOR ENHANCING CELL-FREE METABOLITE PRODUCTION

A version of this chapter by David C. Garcia, Jaime Lorenzo N. Dinglasan, Him Shrestha, Paul E. Abraham, Robert L. Hettich, and Mitchel J. Doktycz, was submitted to Metabolic Engineering Communications and is available as: Garcia, D. C.; Dinglasan J. N.; Shrestha. H.; Abraham, P. E.; Hettich, R. L.; Doktycz M. J. A Lysate Proteome Engineering Strategy for Enhancing Cell-Free Metabolite Production. bioRxiv 2020.

Abstract

Cell-free systems present a significant opportunity to harness the metabolic potential of diverse organisms. Removing the cellular context provides the ability to produce metabolites without the need to maintain cell viability and enables metabolic engineers to explore novel metabolic systems. Crude extracts maintain much of a cell's metabolic capabilities. However, only limited tools are available for engineering the contents of the extracts used for cell-free systems. Thus, our ability to take full advantage of the potential of crude extracts for cell-free metabolic engineering is limited. Here, we employ Multiplex Automated Genomic Engineering (MAGE) to allow selective removal of proteins from crude extracts to specifically direct metabolite production. Specific edits to central metabolism are possible without significantly impacting cell growth. Selective removal of pyruvate degrading enzymes results in engineered crude extracts that are capable of 10-fold increases of pyruvate production when compared to the non-engineered extract. The described approach melds the tools of systems and synthetic biology for the development of cell-free metabolic engineering into a significant platform for both bio-prototyping and bioproduction.

Introduction

Driven by the prospect of biological systems that can be easily manipulated, the application of synthetic biology tools to in vitro environments offers a promising approach to harnessing an organism's rich metabolic potential[107]. Cell-free systems use cytoplasmic components, devoid of genetic material and membranes, as a means of producing complex chemical transformations. While living cells require membranes, growth substrates, and biochemical regulation, in vitro systems sidestep these barriers to manipulation and present an opportunity to explicitly define a system for creating novel proteins and metabolites[108]. In this way, cell-free metabolic engineering (CFME) can use the organism's existing biochemical functions and further combine these capabilities with heterologous pathways to produce chemical precursors, biofuels, and pharmaceuticals.

Efforts to engineer cell-free systems have taken different approaches. Ideally, a CFME system would contain a minimal set of components necessary to carry out a desired biochemical process. Previous approaches employed a defined set of purified enzymes for producing high-yielding chemical conversions and have successfully demonstrated a variety of capabilities including chemical production and protein synthesis[22,109]. Constructing complex, multistep pathways require significant development and upfront costs as utilizing purified proteins at scale remains costly[24]. Further, these purified component systems can exhibit slow catalysis rates, possibly due to the lack of accessory proteins and appropriate protein concentrations capable of improving pathway yield[24]. Nevertheless, long-

running CFME systems that can catalyze multi-step reaction pathways for days have been developed[22].

The use of crude cell extracts presents an alternative approach to CFME. Simple cell lysis and minimal fractionation can be rapidly carried out and result in complex enzyme mixtures for a fraction of the cost of purified components[7,25,91]. Crude extract systems are generally derived from both commonly used cell-free model organisms, such as *E. coli* BL21 Star (DE3), contain a similar biochemistry to the cell of origin, and can serve as both prototyping tools for in vivo metabolic engineering and as bioproduction platforms[8,19,110,111]. Cell-free systems work well for this purpose as extracts can be modularly assembled with those enriched for metabolic pathways in order to produce a specific molecule. Additionally, their compatibility with chemical reactors and ability to consume low-cost feedstocks have popularized them as potential sources for industrial production[9,112,113].

While environmental variables of a cell-free system can easily be manipulated, the proteomic content of the crude extract is more difficult to engineer. Genetic manipulation of a donor strain can substantially impact its growth and function as a bioproduction system[30]. It has been noted previously that simple variations in growth conditions can lead to complex changes in the proteome and significant differences in metabolite flux in the resulting crude extracts[30,114]. Further, specific enzymes can be added or expressed in an extract to further define metabolite production[7,107]. However, removing specific proteins is challenging as gene deletions can affect the growth and global expression of the donor cell. In

particular, deletions to central metabolism can be lethal, which severely limits the ability to direct flux from simple carbon sources. The inability to remove specific pathways from CFME reactions poses a significant constraint and limits the use of crude extracts for bioproduction[7,30]. Tools that allow shaping of the proteome and tailoring of specific pathways will be critical in enabling the use of crude extract systems for metabolic engineering applications.

In this work, we describe the use of genome engineering, specifically MAGE, to enable the removal of particular proteins from crude extracts for CFME. 6xHis-tags are incorporated into proteins that are expected to consume pyruvate and are used for the affinity-based removal of these proteins following cell lysis. Pyruvate, a key connection point in central carbon metabolism, was chosen due to its role linking glycolysis and the Krebs cycle, as well as its relevance as a central precursor for numerous products[90,104]. The use of genome engineering results in a modified proteome capable of producing engineered metabolic phenotypes with minimal impact on the viability of the donor cell (**Figure 4.14**)[115,116]. This general strategy was demonstrated using *E. coli* BL21 Star (DE3) as a chassis due to its prevalent use for cell-free synthetic biology.

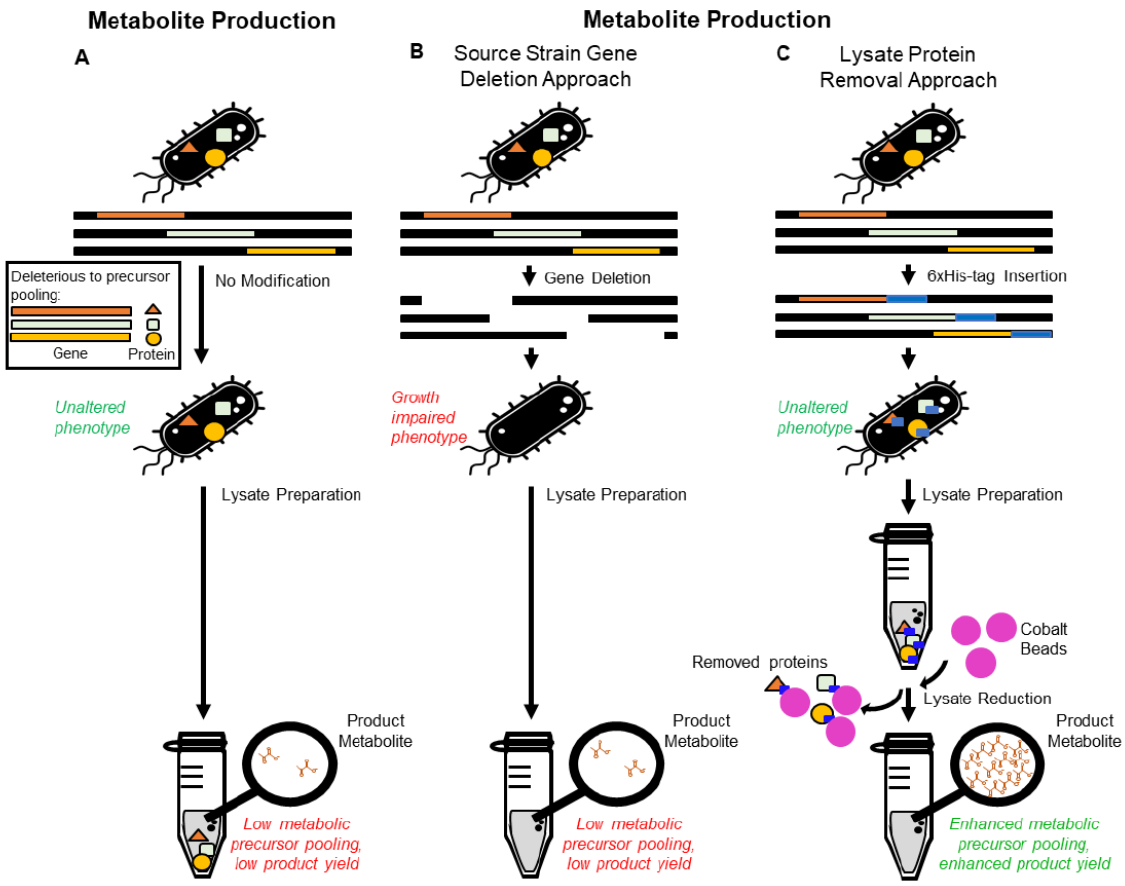

Figure 4.14. Overview of approaches to preparing lysates for cell-free metabolic engineering. A. Complex metabolism present in E. coli lysates harnessed for cell-free metabolite production can compromise central metabolic precursor yields. B & C. Cell-free metabolic engineering approaches seek to reduce lysate complexity in order to redirect carbon flux and pool central metabolic precursors. B. The standard CFME approach reduces lysate complexity by deleting target genes from the source strain, often resulting in growth impaired or lethal phenotypes due to the inability to remove essential genes. This can require multiple design-build-test cycles. C. The new approach involves engineering source strains to endogenously express recognition sequences, such as 6xHis-tags, into target proteins for subsequent removal from lysates through affinity purification, resulting in minimal to no impact on source strain growth and enhanced pooling of specific metabolic products.

Results and Discussion

Pyruvate sits at the biochemical junction of glycolysis and the TCA cycle. It is a key intermediate in producing many food, cosmetic, pharmaceutical and agricultural products whose improved production has been largely unexplored in

cell-free systems[90,104]. In order to create a pyruvate pooling phenotype in an *E. coli* cell-free extract, four proteins were chosen as targets for removal, LdhA, PpsA AceE, and PflB, due to their direct role in consuming pyruvate **(Table 4.6) (Figure 4.15)**[104]. 6xHis tags were fused on either the amino or carboxyl terminus by genetic modification based on an evaluation of literature related to previous purification attempts or crystal structures to find a non-inhibitory location **(Table 4.6)**[117,118].

Table 4.6. Gene and protein information for MAGE targets with a potential effect on pyruvate metabolism.

Gene	Protein	6xHis-Tagged Terminus	MW (kDa)
pflB	Pyruvate formate-lyase	N-Terminal	85
ldhA	D-lactate dehydrogenase	N-Terminal	36.53
ppsA	Phosphoenolpyruvate synthase	N-Terminal	87.43
aceE	Pyruvate dehydrogenase E1	C-Terminal	99.66

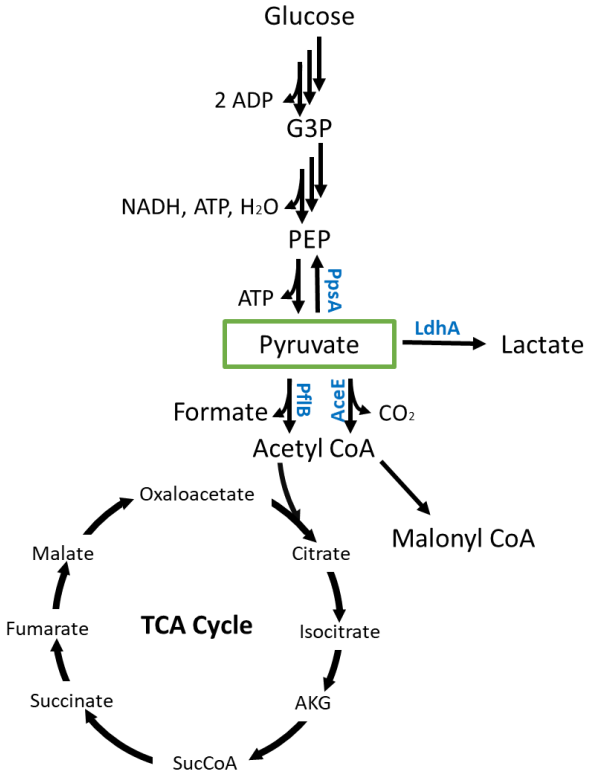

Figure 4.15. Glycolysis and engineered pathway nodes. The numbers 1, 2, 3, and 4 indicate the enzymes pflB, ldhA, ppsA, aceE.

The traditional cell-free protein synthesis chassis *E. coli* BL21 Star(DE3) was chosen as the source strain for creating edited extracts. For genome engineering, the pORTMAGE system was used instead of a genome integrated system due to its potential transferability to multiple donor organisms including *E. coli* BL21 Star(DE3) [119,120]. Additionally, pORTMAGE is curable following genome engineering and relieves the metabolic burden on the cell that can be imparted due to plasmid maintenance[121]. Colony screening was performed using MASC-PCR and further verified using Sanger sequencing **(Figure 4.16 A and B)**[122]. A

total of 5 strains were used for this work. **(Table 4.7)**. The strains included 6xHis-1 (6xHis-*pflB*), 6xHis-2 (6xHis-*pflB-ldhA*), 6xHis-3 (6xHis-*pflB-ldhA-ppsA*), 6xHis-4 (6xHis-*pflB-ldhA-ppsA-aceE*) and 6xHis-*ldhA*. 60 rounds of MAGE were needed to incorporate all four of the tags into the *E. coli* genome **(Figure 4.16 A, B, and C) (Table 4.7)**. This is high when compared to studies producing single nucleotide changes, but in line with other efforts using 6xHis-tags with a genome incorporated MAGE system[109].

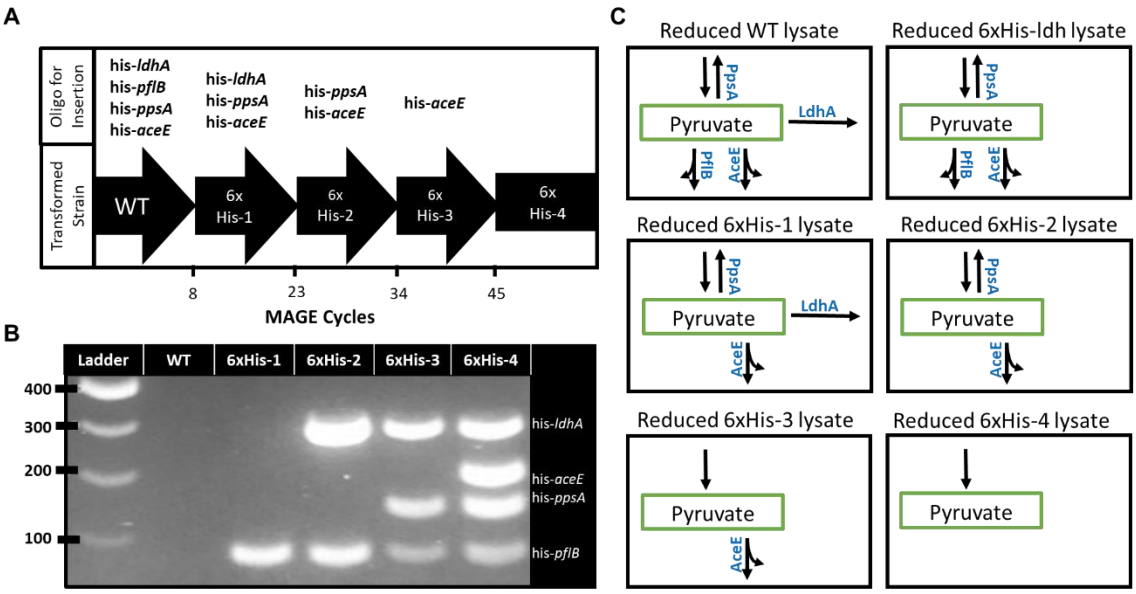

Figure 4.16. Source strain multiplex genome engineering and expected metabolic phenotypes of derived lysates post-reduction. **A.** Strain construction course by MAGE cycling culminating with the 6xHis-4 containing all 4 tags. Each arrow designates the strain being taken through the MAGE process with the oligos used to transform each strain above the arrow. **B.** MASC-PCR results for additive mutations using primers specifically designed for the 6xHis-tagged version of the gene. **C.** Expected metabolic phenotypes present in WT and engineered lysate proteomes after the reduction of lysates derived from all generated strains.

Table 4.7. Strains created for lysate engineering study.

Strain Name	Background	Modified Genes
6xHis-ldh	BL21 Star(DE3)	his-*ldhA,*
6xHis-1	BL21 Star(DE3)	his-pflB
6xHis-2	BL21 Star(DE3)	his-*ldhA,* his-*pflB*
6xHis-3	BL21 Star(DE3)	his-*ldhA,* his-*pflB,* his-*ppsa*
6xHis-4	BL21 Star(DE3)	his-*ldhA,* his-*pflB,* his-*ppsa,* his-*aceE*

After curing each strain of the pORTMAGE plasmid, potential inhibitory effects on growth caused by the expression of tagged proteins were evaluated. Though the presence of the polyhistidine tags has previously been observed to cause growth defects due to the stability of tagged proteins, none of the cells produced for this work showed a significant drop in growth rate (**Appendix Figure 4.19**)[123,124].

The effect of proteome reduction on the extract's metabolic profile was then tested by measurement of pyruvate accumulation in a CFME reaction mix. Early time points show a high accumulation of pyruvate starting at 2 hours and continuing up to 12 hours before the concentration of the metabolite decreases **(Figure 4.17)**. After 8 hours, the reduced 6xHis-4 extract reaches a maximum concentration of 25 mM pyruvate from an initial glucose concentration of 100 mM

(Figure 4.17 A and B). The accumulation of pyruvate at early time points is consistent with a reduction in the proteins capable of consuming pyruvate. In general, the extracts reduced for ldhA show significant increases in pyruvate pooling potentially indicating it is the main in vitro consumer of pyruvate. Interestingly, the effect was not permanent as the pyruvate concentration decreased in the reduced extracts after 12 hours, falling to 6.41 mM for 6xHis-2, 7.2 mM for 6xHis-3, and complete consumption for 6xHis-4 by 24 h **(Figure 4.17)**. For the unreduced strains, extracts not incubated with cobalt beads, little to no pooling of pyruvate occurred from 0 to 8 hours; a result that is corroborated by previous observations of *E. coli* extracts[30]. This is notable as the highest pyruvate concentration between 0 to 8 hours, 4.15 mM in unreduced 6xHis-3, was 5-fold lower than reduced versions of 6xHis-2, 6xHis-3, and 6xHis-4 **(Figure 4.17)**.

The rapid accumulation of pyruvate in the reduced strains after 4 hours leads to a relatively stable concentration in reduced 6xHis-2, 6xHis-3, and 6xHis-4 extracts showing that the reduced lysates are able to maintain an equilibrium between glucose and pyruvate concentrations **(Figure 4.17, Appendix Figure 4.20)**. The rapid consumption of pyruvate following 12 hours for the 6xHis-2, 6xHis-3, and 6xHis-4 extracts may indicate that residual pyruvate consuming machinery maintains activity, but enzymes responsible for feeding PEP or cofactors into the system begin to degrade[125]. Additionally, while the concentration of NAD^+ is bolstered by its exogenous addition during the reaction setup, the elimination of an NADH sink, LdhA, may lead to decreased pooling of pyruvate at later timepoints[7].

These time-dependent changes may indicate that the glycolytic machinery deteriorates faster than the remaining pyruvate degrading pathways. Alternatively, the rapid consumption of the NAD^+ supply could be limiting due to the potential lack of cofactor recycling initiated by the removal of ldhA. Pyruvate consumption experiments performed with the WT and 6xHis-4 lysates show that a significant portion of the pyruvate consuming pathways remain robust after reduction evidencing that the constraint may be due to bottlenecks in upstream glycolysis (**Appendix Figure 4.21**).

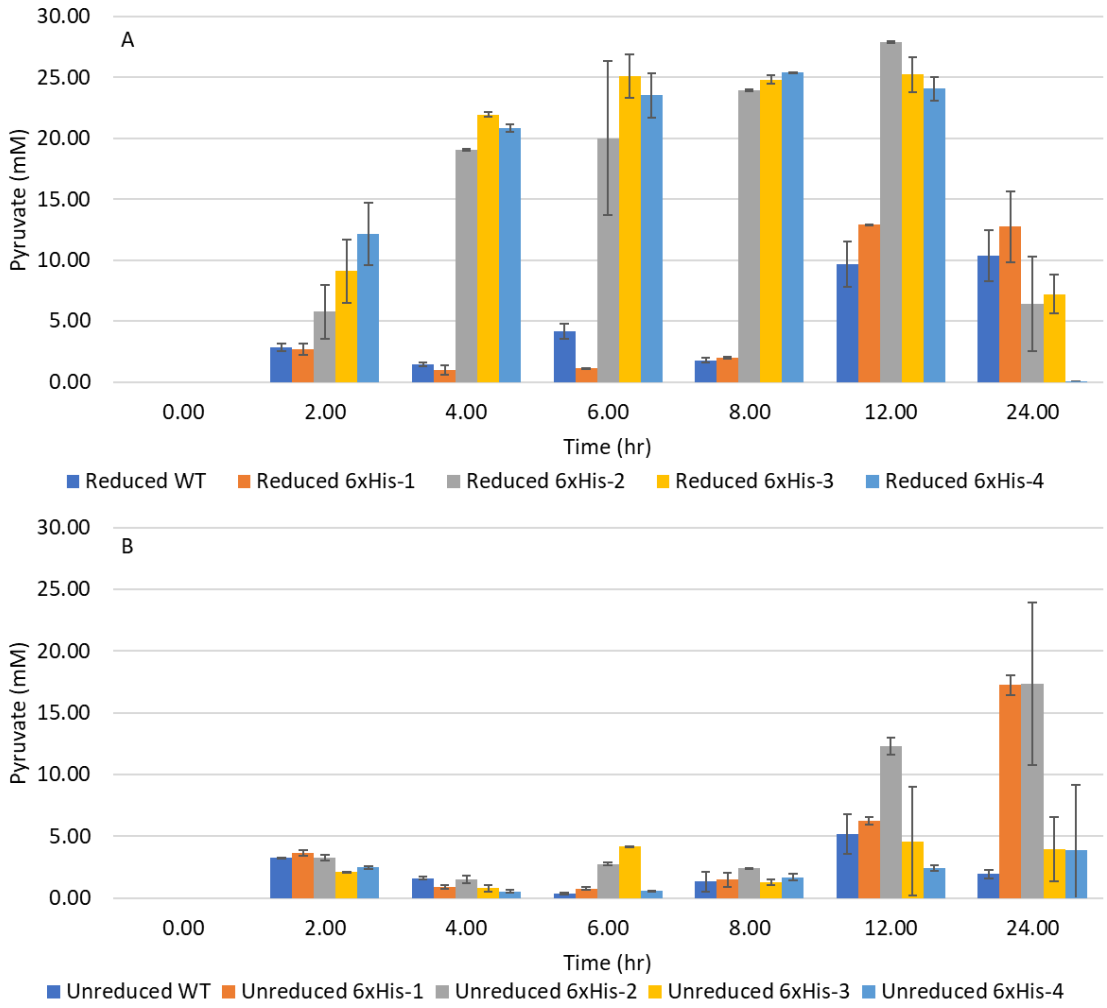

Figure 4.17. Comparisons of pyruvate concentration over time in WT, 6xHis-1, 6xHis-2, 6xHis-3, and 6xHis-4 lysates. Data and standard deviation for the time course reactions were acquired using n=3 biological replicates.

To validate the observed additive effects of successful LdhA and PpsA removals on metabolic output, we generated a strain with a single 6xHis-*ldhA* insertion and compared its extent of pyruvate pooling with the 6xHis-4-reduced lysates **(Appendix Figure 4.22)**. Importantly, the extract showed that while the reduction of LdhA significantly impacts pyruvate accumulation, it was not solely responsible for the effect. At its peak the 6xHis-*ldhA* reduced extract had only 15.73 mM

pyruvate following reduction, falling significantly short of the 27.92 mM, 25.24 mM, and 25.39 mM pyruvate values for 6xHis-2, 6xHis-3, 6xHis-4 mutants, respectively **(Figure 4.17 and Appendix Figure 4.22)**. This result highlights the importance of using multiple 6xHis fusions as well as the potential need for developing cell-free specific flux balance models for target selection.

To further understand the pyruvate pooling trends resulting from the use of engineered lysates, differences between the unreduced and reduced proteomes of each of the lysates were evaluated by mass spectrometry **(Figure 4.18)**. As expected for the WT extract, the reduction process did not significantly impact the presence of the targeted proteins when comparing the peptide abundance values of unreduced and reduced versions **(Figure 4.18 A)**. On the other hand, a substantial 3-to 4-fold decrease in relative LdhA abundance was evident in reduced 6xHis-2, 6xHis-3, and 6xHis-4 lysates compared to their unreduced counterparts **(Figure 4.18 C, D, and E)**, which is consistent with the significant increase in these extracts' ability to accumulate pyruvate **(Figure 4.16)**. PpsA was also significantly pulled-down from only the 6xHis-3 and 6xHis-4 lysates. Comparing the metabolic output of these lysates to that of the WT, we observed a 5- to 10-fold increase in pyruvate accumulation at earlier time points **(Figure 4.17)**.

From the mass spectrometry-based proteomics profiling, it is evident that 6xHis-tagged LdhA and PpsA could be removed from lysates, while significant removal of 6xHis-tagged PflB was not successful **(Figure 4.18 B to E)**. Although the decrease in PflB levels between the unreduced and reduced 6xHis-4 lysates met

the significance threshold (pval < 0.05), the change was only a 1.81-fold reduction compared to the significant decreases in LdhA and PpsA following lysate reduction **(Figure 4.18 E)**. AceE was not observed to be pulled down after the purification of 6xHis-tagged proteins from the 6xHis-4 lysate. These findings are consistent with metabolic output data in that the reduced 6xHis-1 and 6xHis-4 lysates seem to pool as much pyruvate as the reduced WT and 6xHis-3 lysates, respectively **(Figure 4.17)**.

In their native context, such as in crude cell extracts, both PflB and AceE are known to interact with other proteins. PflB binds to the cytoplasmic membrane protein formate channel A, FocA, to facilitate the translocation of formate and this strong association in *E. coli* lysates has been demonstrated by co-purification experiments[126]. AceE is the pyruvate-binding protein of the pyruvate dehydrogenase complex (PDHC), a stable multi-enzyme complex comprised of two additional components, dihydrolipoyl transacetylase and dihydrolipoyl dehydrogenase[127]. Individually, overexpressed PflB and AceE have been successfully isolated from their interaction partners through 6xHis-tag purification in the past. However, under native expression and structural context, the use of 6xHis-tags may not be effective for purifying complexed proteins as shorter tags may not destabilize complexes[117,126]. We reason that the formation of these protein-protein interactions may be restricting the accessibility of the target proteins' termini to the affinity resin, and therefore a larger protein tag that positions itself outside of a complex may be a more favorable alternative[128]. Alternatively,

the 6xHis-tags of PflB and AceE are efficiently exposed to the resin, but binding is sterically hindered by the large protein complexes. In this case, the application of protein tags known to efficiently purify intact protein complexes may improve the reduction method[129].

We further analyzed proteomic changes in the unreduced and reduced extracts to determine whether the reduction method resulted in the removal of off-target proteins in reduced lysates. Importantly, the process of reducing the proteome did not seem to significantly impact proteins in central metabolism outside of those targeted by the reduction process. When comparing the reduced and unreduced WT lysates, in the 58 proteins with a greater than 4-fold reduction, none were connected to central metabolism outside of roles in membrane transport (Supplemental Data 1). Future efforts will seek to minimize off-target effects by using degradation instead of 6xHis tags in order to improve the general applicability of lysate proteome engineering.

Targeted reduction of a lysate proteome enables a rapid means to manipulate central metabolism without the possible drawback of cultivating “sick” organisms as often results from traditional, in vivo metabolic engineering efforts. More broadly, this new approach to engineering in vitro metabolism yields metabolic states that are not traditionally possible in living organisms and sets important groundwork for further developments in cell-free metabolic engineering technology. The pORTMAGE system offers the potential for extension of this engineering strategy to other, non-model cell-free chassis. Future improvements

to lysate proteome engineering could also entail the use of multiplex genome engineering methods that are amenable to the insertion of larger tags[120,130,131]. To further advance the reduction of the lysate's proteome, orthogonal protein degradation systems could be employed wherein proteins are genomically tagged and degraded in a cell-free extract using an exogenous protease. The *mf-lon* protease system serves this function. This protease targets a 27 amino acid long peptide and may overcome issues associated with targeting protein complexes[132,133].

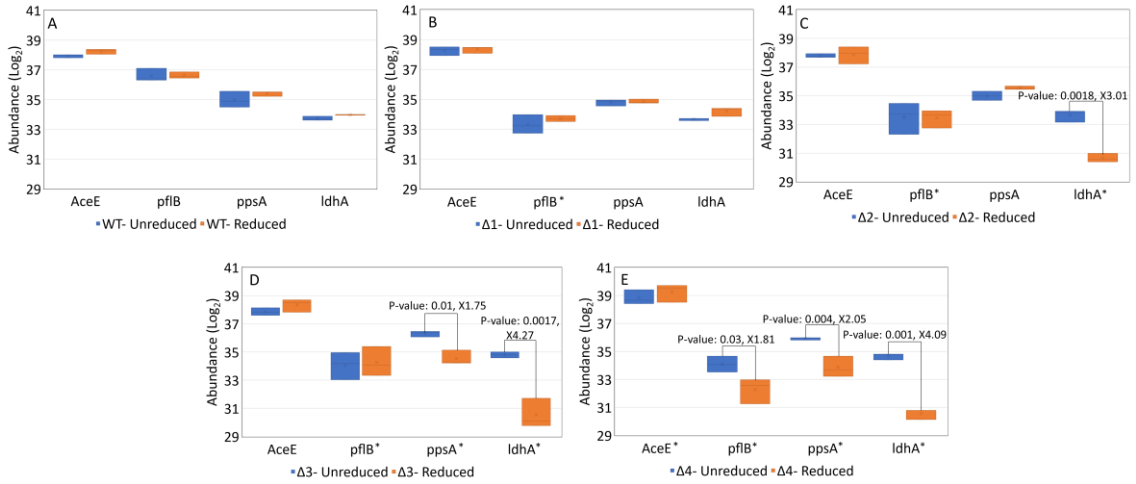

Figure 4.18. Proteomic analysis of an unreduced cell-free extract (blue), and a reduced cell-free extract (orange). Significant fold-changes in protein concentration when comparing the reduced to the unreduced extract are denoted by p-value and fold change reduction in concentration of the protein above a bracket. Panels A, B, C, D, and E denote the WT, 6xHis-1, 6xHis-2, 6xHis-3, and D. 6xHis-4 strains, respectively. Asterisks indicate proteins targeted for removal in the reduced strain.

Methods

Generation and Validation of Genome Engineered Strains using MAGE

All Multiplex allele-specific PCR (MASC-PCR), Sanger Sequencing oligos, and recombineering oligos were created manually and ordered from IDT with standard purification. Each targeting oligo incorporated four phosphorothioated bases on the 5′ terminus. An 18-base CACCATCACCATCACCAT sequence was used to add the His-tag and directed at either the N- or C-terminus based on previous literature or crystal structure analysis[117,118]. The pORTMAGE protocol used in this study followed previous work with the exception that growth was carried out in 6 mL of Luria-Bertani-Lennox (lbl) cultures in glass tubes with 100 mg/mL of carbenicillin, recovery was performed in 3 mL of terrific broth with a 1-hour incubation time prior to adding 3 mL of lbl-carb for outgrowth[134]. Given the significant time required to find accumulated mutations in a single strain, the additive mutations were started from previously found mutations such that Δ1 was used to create Δ2 and so on as per the protocols used in previous studies[109]. After every 8-12 cycles of MAGE, 30-60 colonies were screened for genome edits using MASC-PCR as detailed previously. Allelic genotyping was performed using standard primers designed to flank both modified genes. Amplicons were Sanger sequenced to validate the insertion of the His-tag sequence. Primer sequences used in this study are listed in (**Appendix Table 4.8**) and (**Appendix Table 4.9**).

Cell-Free Extract Preparation Protocol

Following plasmid curing, the cell extracts were prepared from *E. coli* BL21 Star (DE3) grown at 37 °C in 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). Cell extracts were prepared by harvesting 50-mL cultures grown in baffled Erlenmeyer flasks to an OD600 of 4.5 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. The pellets were weighed, flash-frozen and stored at –80 °C. Extracts were prepared by thawing and resuspending the cells in 0.8 mL of S30 buffer per mg of cell wet weight. The resuspension was lysed using 530 joules per mL of suspension at 50% tip amplitude with ice water cooling. Following sonication, tubes of cell extract were centrifuged twice at 21,100 x g for 10 minutes at 4 °C, aliquoted, frozen with liquid nitrogen, and stored at –80 °C.

Cell-Free Extract Reductions

Cell extracts were reduced by adding one volume of cell extract to 0.2X volume of ice-cold HisPur™ Cobalt Resin (ThermoFisher Scientific) suspension in 1.5 mL microcentrifuge tubes. Prior to the addition of lysate, HisPur™ Cobalt Resin was washed 2X with 500 µL S30 buffer and incubated with 10 mM imidazole buffer (pH 4.5; 10 mM imidazole, 50 mM monopotassium chloride, 300 mM NaCl). Lysate-resin mixtures were incubated for 1 hour at 4˚C under shaking conditions (800 rpm) to ensure the suspension of the resin particles in the extracts then centrifuged at

14,000 x g for 30 seconds. Supernatants were aliquoted, flash-frozen, and stored at -80˚C until used. His-tagged proteins were eluted from the HisPur™ Cobalt Resin by suspending the resin in 50 µL elution buffer (pH 4.5; 250 mM Imidazole, 50 mM monosodium phosphate, 300 mM NaCl) for 30 minutes at 4˚C under shaking conditions (800 rpm). The eluate was obtained for proteomic quantification by spinning down the suspension at 14,000 x g for 30 seconds and collecting the supernatant.

CFME Reaction Set-up

Glucose consumption reactions were carried out at 37 °C in 50 µL volumes using a solution of 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. Similarly, pyruvate fed reactions were carried out using the conditions with the exception of 25 mM pyruvate being used in place of glucose. Extracts were added to a final protein concentration of 4.5 mg mL-1. Each reaction was quenched by the addition of 50 µL of 5% TCA. The supernatant was centrifuged at 11,000 x g for 5 minutes and directly used for analytical measurements.

Proteomics Sample Preparation

Cell extracts were solubilized in 200 µL of 4% SDS in 100 mM Tris buffer, pH 8.0. Trichloroacetic acid was added to achieve a concentration of 20% (wt/vol). Samples were vortexed and incubated at 4°C for 2 h followed by 10 min at -80°C. Samples were then thawed on ice prior to centrifugation (~21,000g) for 10 min at

4°C to pellet precipitated proteins from the detergent and solutes. The supernatant was discarded, and samples were washed with 1 mL of ice-cold acetone. Pelleted proteins were then air-dried and resuspended in 100 µL of 8 M urea in 100 mM Tris buffer, pH 8.0. Proteins were reduced with 10mM dithiothreitol incubated for 30 min and alkylated with 30mM iodoacetamide for 15 min in dark at room temperature. Proteins were digested with two separate and sequential aliquots of sequencing grade trypsin (Promega) of 1 µg. As 8 M urea inhibits trypsin, samples were diluted to 4 M urea for a 3 hours digestion, followed by dilution to 2 M urea for overnight digestion. Samples were then adjusted to 0.1% trifluoroacetic acid and then desalted on Pierce peptide desalting spin columns (Thermo Scientific) as per the manufacturer instructions. Samples were vacuum-dried with a SpeedVac (Thermo Scientific) and then resuspended in 50 µL of 0.1% formic acid. Peptide concentrations were then measured using nanodrop (Thermo Scientific) and 2 µg of each sample was used for LC-MS measurement.

LC-MS/MS analysis

All samples were analyzed on a Q Exactive Plus mass spectrometer (Thermo Scientific) coupled with an automated Vanquish UHPLC system (Thermo Scientific). Peptides were separated on a triphasic precolumn (RP-SCX-RP; 100 µm inner diameter and 15 cm total length) coupled to an in-house-pulled nanospray emitter of 75 µm inner diameter packed with 25 cm of 1.7 µm of Kinetex C18 resin (Phenomenex). For each sample, a single 2µg injection of peptides was loaded and analyzed across a salt cut of ammonium acetate (500mM) followed by

a 210 min split-flow (300nL/min) organic gradient, wash, and re-equilibration: 0% to 2% solvent B over 27 min, 2% to 25% solvent B over 148 min, 25% to 50% solvent B over 10 min, 50% to 0% solvent B over 10 min, hold at 0% solvent B for 15 min. MS data were acquired with the Thermo Xcalibur software using the top 10 data-dependent acquisition.

Proteome Database Search

All MS/MS spectra collected were processed in Proteome Discoverer, version 2.3 with MS Amanda and Percolator. Spectral data were searched against the most recent *E. coli* reference proteome database from UniProt to which mutated sequences and common laboratory contaminants were appended. The following parameters were set up in MS Amanda to derive fully tryptic peptides: MS1 tolerance = 5 ppm; MS2 tolerance = 0.02 Da; missed cleavages = 2; Carbamidomethyl (C, + 57.021 Da) as static modification; and oxidation (M, + 15.995 Da) as dynamic modifications. The Percolator FDR threshold was set to 1% at the peptide-spectrum match and peptide levels. FDR-controlled peptides were then quantified according to the chromatographic area-under-the-curve and mapped to their respective proteins. Areas were summed to estimate protein-level abundance.

Proteomic Data Analysis

For differential abundance analysis of proteins, the protein table was exported from Proteome Discoverer. Proteins were filtered to remove stochastic sampling. All proteins present in 2 out of 3 biological replicates in any condition were

considered valid for quantitative analysis. Data was log_2 transformed, LOESS normalized between the biological replicates and mean-centered across all the conditions. Missing data were imputed by random numbers drawn from a normal distribution (width = 0.3 and downshift = 2.8 using Perseus software (http://www.perseus-framework.org). The resulting matrix was subjected to ANOVA and a post-hoc TukeyHSD test to assess protein abundance differences between the different experimental groups. The statistical analyses were done using an in-house developed R script.

Glucose and pyruvate measurements

Pyruvate and glucose measurements were performed using high-performance liquid chromatography (HPLC). An Agilent 1260 equipped with an Aminex HPX 87-H column (Bio-Rad, Hercules, CA) and a diode array UV-visible detector (Agilent, Santa Clara, CA) reading at 191 nm. Analytes were eluted with isocratic 5 mM sulfuric acid at a flow rate of 0.55 mL min–1 at 35 °C for 25 min. Pyruvate and glucose standards were used for sample quantification using linear interpolation of external standard curves.

Conclusion

Shown in this work was the use of genome engineering to affect in vivo protein modifications that lead to directed metabolic activity in the derived lysates. This novel cell-free metabolic engineering strategy allows targeted removal of enzymes, thus enabling the production of metabolites from simple precursors using rapidly prepared crude extracts[135–138]. The ability to extract pyruvate degrading enzymes, leading to unconventional metabolic states, was engineered and shown to be capable of pooling pyruvate for a significant period of time. The ability to direct metabolic flux in cell-free systems and create proteomes untenable to living cells was demonstrated. The flexibility of CFME systems highlights the significant value they hold as novel bioproduction platforms. The advances made in this work can be extended to design molecule specific donor strains for natural product biosynthesis, such as for polyketides or carbohydrates, through the removal of defined inhibitory reactions. The ability to reduce specific components of crude lysates allows for more complex reaction networks to be employed in the development of CFME bioproduction platforms. As CFME begins to tackle new challenges related to antibiotic, fuel, and materials production, innovative engineering tools and techniques will be crucial to advancing the scope and adoption of cell-free biological production.

Appendix

Table 4.8. MAGE oligos use for this study. Phosphorothioated bases are noted with asterisks.

Primer	Sequence
Pfl	a*a*t*a*aaaaatccacttaagaaggtaggtgttacatgCACcatCACcatCACCATtccgagcttaatgaaaagttagccacagcctgggaa
Ldh	t*a*a*a*tgtgattcaacatcactggagaaagtcttatgCACcatCACcatCACCATaaactcgccgtttatagcacaaaacagtacgacaag
Ppsa	c*a*a*a*ccgttcatttatcacaaaaggattgttcgatgCACcatCACcatCACCATtccaacaatggctcgtcaccgctggtgctttggtat
Pdh	a*c*t*c*aacgttattagatagataaggaataacccatgCACcatCACcatCACCATtcagaacgtttcccaaatgacgtggatccgatcgaa

Table 4.9. MASC-PCR oligos used for this study.

Primer	Sequence
Pfl F	GCCAGCCAGGAAGGACTCGTCACCCTCG
Pfl R	GCAGTAAATAAAAAATCCACTTAAGAAGGTAGGTGTTACATGC
Ldh F	CAGCGTCATCATCATACCGATGGC
Ldh R	CTTAAATGTGATTCAACATCACTGGAGAAAGTCTTATGC
Ppsa F	GCTGGTTTACGCCGCTTTGGTCC
Ppsa R	ACCGTTCATTTATCACAAAAGGATTGTTCGATGC
Pdh F	TGGCCTTTATCGAAGAAATTTTGCTCGACAG
Pdh R	ATCCACGTCATTTGGGAAACGTTCTGAA

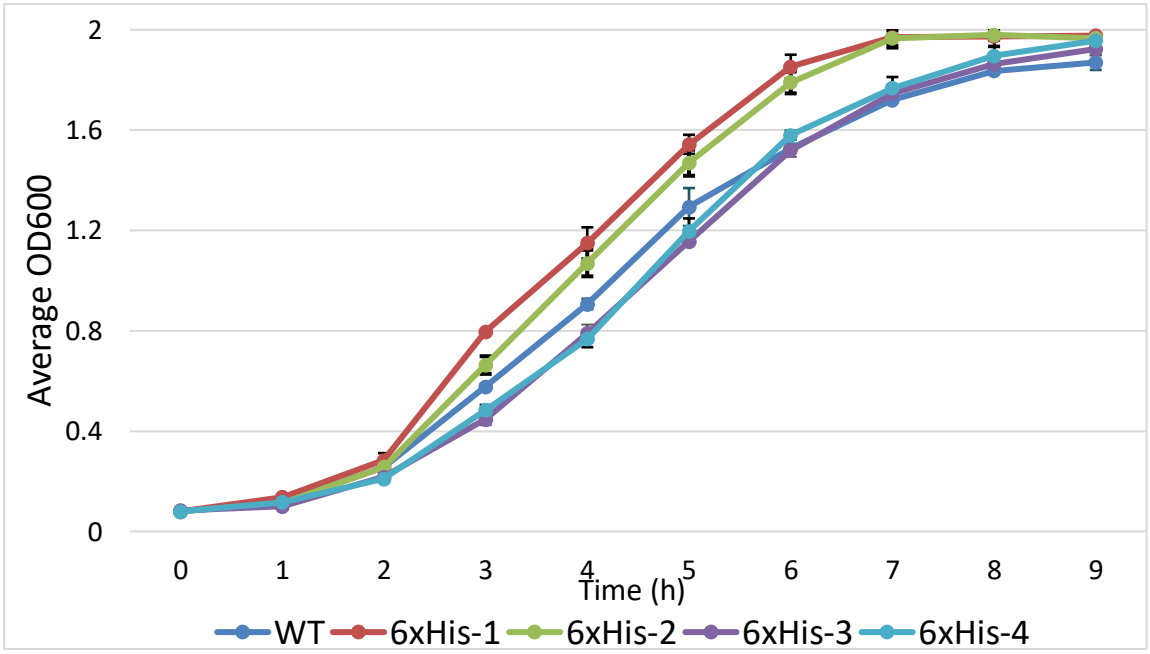

Figure 4.19. Growth rate and terminal OD600 was measured using for the four main strains produced for this study.

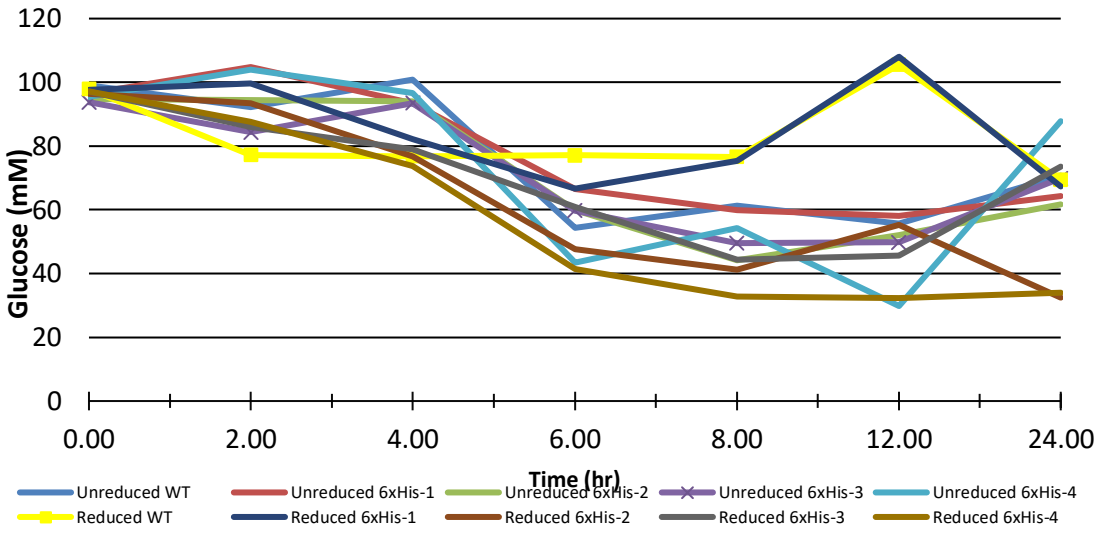

Figure 4.20. Glucose measured over a 24-hour time period.

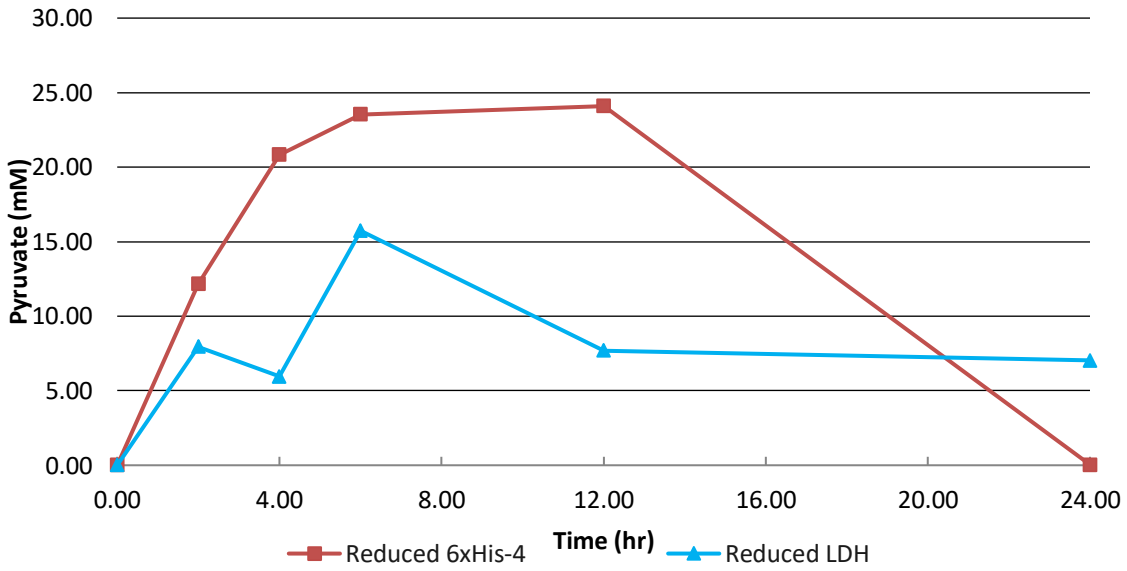

Figure 4.21. Comparison of pyruvate concentration changes when fed 100 mM glucose for an extract with a single reduction of ldhA to an extract reduced for all 4 proteins *his*-ldhA, *his*-pflB, *his*-ppsa, *his*-aceE.

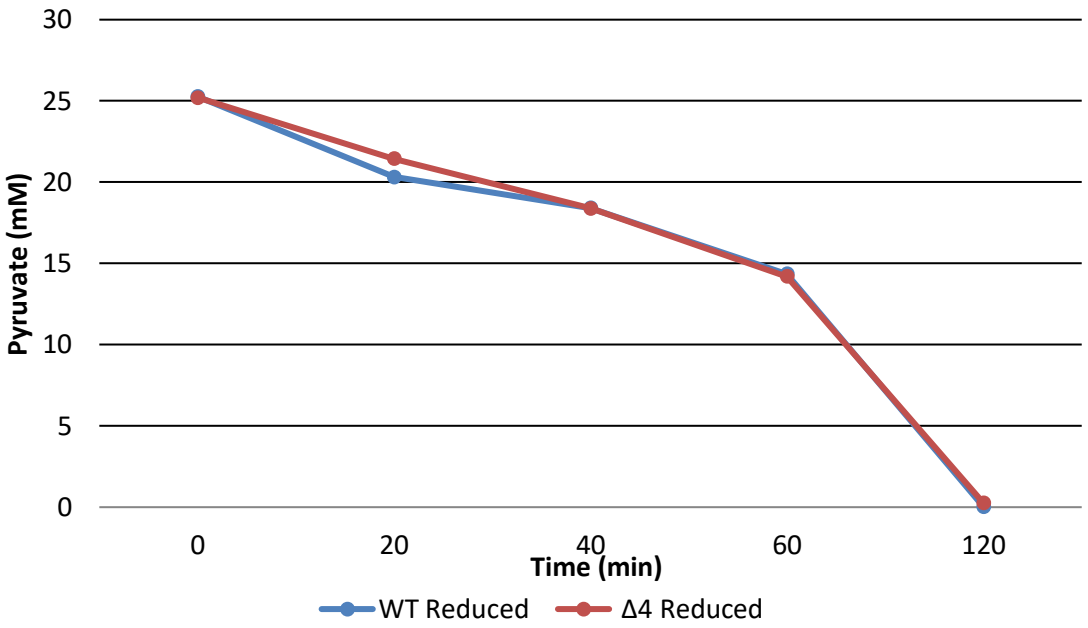

Figure 4.22. Comparison of pyruvate consumption using reduced WT and Δ4 extracts. All reaction conditions were exactly as those used for the glucose consuming reactions aside from the replacement of glucose with 25 mM pyruvate.

Chapter 5

Conclusions

Since their inception, cell-free systems have found uses as powerful research tools in a variety of applications due to their ability to emulate a living system. What once helped decipher the genetic code is now being applied to biological production, protein characterization, and the development of synthetic cells[139]. As the challenges of synthetic biology and biological research continue to expand, so too will the technology of cell-free systems need to develop in order to keep pace and fulfill future needs. The work presented in this dissertation provides tools and applications to expand the use of cell-free systems to showcase their application towards solving complex biological problems.

Specifically, this work focus on the development of cell-free systems as conduits for the exploration of cellular metabolism and in vitro biological production from microorganisms. Due to the advent of next-generation sequencing and genome mining techniques, our continuously expanding repositories of biological information have been left with few tools for rapid and empirical verification. These datasets contain significant amounts of information regarding the function, stability, and metabolic output of the microbial world. Whereas traditional in vivo systems can be limited by cytotoxicity and the metabolic burdens of living cells, cell-free systems are only limited by an understanding of their chemical processes. The scalable nature of cell-free systems present an excellent opportunity to both elucidate the function of microbial dark matter through pathway identification and to harness the metabolic potential of microbial organisms through cell-free metabolic engineering.

To the ends described above, I published three papers, presented here as chapters, exploring uncharacterized metabolic pathways, analyzing the effects of growth conditions on extracts, and developing tools to engineer cell-free metabolism[25,30,140]. A key aspect of these projects is the role they share in taking cell-free biology and further expanding our understanding of its capabilities to facilitate its use. Each work, therefore, reflects an improved understanding of the role of cell-free systems as research and industrial tools.

Chapter one attempted to express metabolic pathways outside of their native context to showcase the utility of crude cell-free extracts as a verification tool for computationally predicated metabolic pathways. These efforts allowed us to elucidate the metabolic pathways used by *Pantoea sp.* YR 343 to produce IAA. The key feature of our approach is the combination of computational and empirical tools to both rapidly predict and verify the activity of a given pathway. This platform ultimately allowed us to show that *Pantoea sp.* YR 343 was capable of using multiple enzymatic combinations to produce IAA. The insights learned allows for more conventional tools such as genome and metabolic engineering to be used without a significant fear of wasted resources and time.

Chapter two seeks to improve pyruvate production in cell-free systems, by measuring the impacts modifications to media and preparation conditions can have on protein content and showed how these changes can be leveraged to improve biosynthetic processes in vitro. The effects of four growth conditions for a crude S30 *E. coli* lysate were assessed with shotgun proteomics and analytical

measurements. Our use of proteomic techniques helped to validate the metabolic measurements made from CFME reactions. Ultimately, we showed that the proteome of a cell-free extract can have a significant impact on the performance and function of a cell-free extract as each extract both pooled and consumed pyruvate at significantly different levels. These experiments furthered our interest in CFME as the Y-TE extract while being capable of pooling a large amount of pyruvate was not doing so in stoichiometric quantities. This indicated that a significant portion of our feeding substrate was being consumed by side reactions. A problem that would need to be solved before CFME systems could have a significant impact as bioproduction platforms.

Chapter three sought to help resolve some of the problems associated with CFME by designing a cell-free specific genome engineering system. While traditional tools would simply remove proteins deleterious to the production of a specific metabolite and thereby risk a cytotoxic phenotype, the system we designed for this work alleviated this concern by performing the deletion process after lysis. Reducing the cell-free proteome enables complex reaction networks to be modified for the development of CFME. We successfully engineered an unconventional metabolism able to produce pyruvate and showed an important proof of principle for creating proteomes untenable to living cells.

Future efforts will seek to expand the throughput and efficiency of the work described in this dissertation. While ligand docking using Autodock Vina is useful for rapidly predicting protein function, more powerful computational tools such a

molecular dynamics, though slow outside of large computational clusters, would significantly improve the accuracy of the protein-ligand interactions. The use of liquid handling systems could also significantly increase the rate at which predicted enzymes could be tested and pathways verified. One problem seen in the development of our reduced cell-free metabolic engineering extracts was the incomplete degradation of the tagged proteins. A promising avenue for a more complete reduction is the use of an orthogonal protein degradation system such as that of the mf-Lon protease. Instead of removing the proteins from the extract, degradation tags would be introduced into target proteins and degraded after lysis of the cell[133]. Though at the moment, such large tags can only be added in multiplex to organisms with natural competence like *Vibrio natriegens*, the process could also be replicated in *E. coli* as advances to MAGE are released.

Cell-free systems derive their function from a complex set of biochemistries removed from the context of cellular regulation. Free from membranes and endogenous genetic material, the relatively static proteome becomes open to more dramatic changes to its environment. This enables cell-free systems to elucidate the functions of complex chemical networks or generate molecules using unconventional metabolic states. As the need for scalable biological tools continues to expand, these characteristics will further incentivize the study and use of cell-free systems. Continuing to develop these powerful tools will enable scientists and engineers to create novel biological production platforms cheaply and sustainably.

REFERENCES

(1) Nirenberg, M. W., and Matthaei, J. H. (1961) The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. 47*, 1588–1602.

(2) Lamborg, M. R., and Zamecnik, P. C. (1960) Amino acid incorporation into protein by extracts of E. coli. *Biochim. Biophys. Acta 42*, 206–211.

(3) ZAMECNIK, P. C., and FRANTZ, I. D. (1948) Incorporation *in vitro* of radioactive carbon from carboxyl-labeled dl-alanine and glycine into proteins of normal and malignant rat livers. *J. Biol. Chem. 175*, 299–314.

(4) KELLER, E. B., and LITTLEFIELD, J. W. (1957) Incorporation of C14-amino acids into ribonucleoprotein particles from the Ehrlich mouse ascites tumor. *J. Biol. Chem. 224*, 13–30.

(5) Carlson, E. D., Gan, R., Hodgman, E. C., and Jewett, M. C. (2014) Cell-Free Protein Synthesis : Applications Come of Age *30*, 1185–1194.

(6) Wu, Y. Y., Culler, S., Khandurina, J., Van Dien, S., and Murray, R. M. (2015) Prototyping 1,4-butanediol (BDO) biosynthesis pathway in a cell-free transcription-translation (TX-TL) system. *bioRxiv* 017814.

(7) Kay, J. E., and Jewett, M. C. (2015) Lysate of engineered Escherichia coli supports high-level conversion of glucose to 2,3-butanediol. *Metab. Eng. 32*, 133–142.

(8) Hurst, G. B., Asano, K. G., Doktycz, C. J., Consoli, E. J., Doktycz, W. L., Foster, C. M., Morrell-Falvey, J. L., Standaert, R. F., and Doktycz, M. J. (2017) Proteomics-Based Tools for Evaluation of Cell-Free Protein Synthesis. *Anal. Chem.* acs.analchem.7b02555.

(9) Timm, A. C., Shankles, P. G., Foster, C. M., Doktycz, M. J., and Retterer, S. T. (2016) Toward Microfluidic Reactors for Cell-Free Protein Synthesis at the Point-of-Care. *Small 12*, 810–817.

(10) Spirin, A., Baranov, V., Ryabova, L., Ovodov, S., and Alakhov, Y. (1988) A continuous cell-free translation system capable of producing polypeptides in high yield. *Science (80-.). 242*, 1162–1164.

(11) Chen, H.-Z., and Zubay, G. (1983) [44] Prokaryotic coupled transcription — translation, pp 674–690.

(12) Shimizu, Y., Inoue, a, Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001) Cell-free translation reconstituted with purified components. *Nat. Biotechnol. 19*, 751–755.

(13) Villarreal, F., Contreras-llano, L. E., Chavez, M., Ding, Y., Fan, J., Pan, T., and Tan, C. (2017) assembly of pure translation machinery.

(14) Lavickova, B., and Maerkl, S. J. (2019) A Simple, Robust, and Low-Cost

Method to Produce the PURE Cell-Free System. *ACS Synth. Biol. 8*, 455–462.

(15) Wang, H. H., Huang, P., Xu, G., Haas, W., and Marblestone, A. SP-Multiplexed in vivo His-tagging of enzyme pathways for *in vitro* single-pot multi-enzyme catalysis.

(16) Ezure, T., Suzuki, T., Shikata, M., Ito, M., Ando, E., Utsumi, T., Nishimura, O., and Tsunasawa, S. (2010) Development of an Insect Cell-Free System. *Curr. Pharm. Biotechnol. 11*, 279–284.

(17) Martin, R. W., Majewska, N. I., Chen, C. X., Albanetti, T. E., Jimenez, R. B. C., Schmelzer, A. E., Jewett, M. C., and Roy, V. (2017) Development of a CHO-Based Cell-Free Platform for Synthesis of Active Monoclonal Antibodies. *ACS Synth. Biol. 6*, 1370–1379.

(18) Harbers, M. (2014) Wheat germ systems for cell-free protein expression. *FEBS Lett. 588*, 2762–2773.

(19) Kay, J. E., and Jewett, M. C. (2015) Lysate of engineered Escherichia coli supports high-level conversion of glucose to 2,3-butanediol. *Metab. Eng. 32*, 133–142.

(20) Karim, A. S., and Jewett, M. C. (2016) A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. *Metab. Eng. 36*, 116–126.

(21) Zaslavsky, L., Ciufo, S., Fedorov, B., and Tatusova, T. (2016) Clustering analysis of proteins from microbial genomes at multiple levels of resolution. *BMC Bioinformatics 17*.

(22) Opgenorth, P. H., Korman, T. P., and Bowie, J. U. (2016) A synthetic biochemistry module for production of bio-based chemicals from glucose. *Nat. Chem. Biol. 12*, 1–4.

(23) Shin, Jonghyeon, V. N. (2011) An E. coli Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. *ACS Synth. Biol.*

(24) Korman, T. P., Opgenorth, P. H., and Bowie, J. U. (2017) A synthetic biochemistry platform for cell free production of monoterpenes from glucose. *Nat. Commun. 8*, 1–8.

(25) C. Garcia, D., Cheng, X., L. Land, M., F. Standaert, R., L. Morrell-Falvey, J., and J. Doktycz, M. (2019) Computationally Guided Discovery and Experimental Validation of Indole-3-acetic Acid Synthesis Pathways. *ACS Chem. Biol. 14*, 2867–2875.

(26) Dudley, Q. M., Karim, A. S., and Jewett, M. C. (2015) Cell-free metabolic engineering: Biomanufacturing beyond the cell. *Biotechnol. J. 10*, 69–82.

(27) Casini, A., Chang, F. Y., Eluere, R., King, A. M., Young, E. M., Dudley, Q. M., Karim, A., Pratt, K., Bristol, C., Forget, A., Ghodasara, A., Warden-Rothman, R., Gan, R., Cristofaro, A., Borujeni, A. E., Ryu, M. H., Li, J., Kwon, Y. C., Wang, H., Tatsis, E., Rodriguez-Lopez, C., O'Connor, S., Medema, M. H., Fischbach, M. A.,

Jewett, M. C., Voigt, C., and Gordon, D. B. (2018) A Pressure Test to Make 10 Molecules in 90 Days: External Evaluation of Methods to Engineer Biology. *J. Am. Chem. Soc. 140*, 4302–4316.

(28) Bothfeld, W., Kapov, G., and Tyo, K. E. J. (2017) A Glucose-Sensing Toggle Switch for Autonomous, High Productivity Genetic Control. *ACS Synth. Biol. 6*, 1296–1304.

(29) C. Garcia, D., Cheng, X., L. Land, M., F. Standaert, R., L. Morrell-Falvey, J., and J. Doktycz, M. (2019) Computationally Guided Discovery and Experimental Validation of Indole-3-acetic Acid Synthesis Pathways. *ACS Chem. Biol. 14*, 2867–2875.

(30) Garcia, D. C., Mohr, B. P., Dovgan, J. T., Hurst, G. B., Standaert, R. F., and Doktycz, M. J. (2018) Elucidating the potential of crude cell extracts for producing pyruvate from glucose. *Synth. Biol. 3*.

(31) Campo, A. M., Bodea, S., Hamer, H. A., Marks, J. A., Haiser, H. J., Turnbaugh, P. J., and Balskus, E. P. (2015) Characterization and Detection of a Widely Distributed Gene Cluster That Predicts Anaerobic Choline Utilization by Human Gut Bacteria *6*, 1–12.

(32) Tan, G.-Y., Deng, Z., and Liu, T. (2016) Recent advances in the elucidation of enzymatic function in natural product biosynthesis. *F1000Research 4*, 1399.

(33) Gerlt, J. A., Allen, K. N., Almo, S. C., Armstrong, R. N., Babbitt, P. C., Cronan, J. E., Dunaway-Mariano, D., Imker, H. J., Jacobson, M. P., Minor, W., Poulter, C. D., Raushel, F. M., Sali, A., Shoichet, B. K., and Sweedler, J. V. (2011) The Enzyme Function Initiative. *Biochemistry 50*, 9950–9962.

(34) Koskinen, P., Törönen, P., Nokso-Koivisto, J., and Holm, L. (2015) PANNZER: High-throughput functional annotation of uncharacterized proteins in an error-prone environment. *Bioinformatics 31*, 1544–1552.

(35) Lamontagne, J., Béland, M., Forest, A., Côté-Martin, A., Nassif, N., Tomaki, F., Moriyón, I., Moreno, E., and Paramithiotis, E. (2010) Proteomics-based confirmation of protein expression and correction of annotation errors in the Brucella abortus genome. *BMC Genomics 11*.

(36) Lewis, K., Epstein, S., D'Onofrio, A., and Ling, L. L. (2010) Uncultured microorganisms as a source of secondary metabolites. *J. Antibiot. (Tokyo). 63*, 468–476.

(37) Cobb, R. E., Wang, Y., and Zhao, H. (2015) High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. *ACS Synth. Biol. 4*, 723–728.

(38) Tripathi, S. A., Olson, D. G., Argyros, D. A., Miller, B. B., Barrett, T. F., Murphy, D. M., McCool, J. D., Warner, A. K., Rajgarhia, V. B., Lynd, L. R., Hogsett, D. A., and Caiazza, N. C. (2010) Development of pyrF-Based genetic system for targeted gene deletion in clostridium thermocellum and creation of a pta mutant. *Appl.*

Environ. Microbiol. 76, 6591–6599.

(39) Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E. J., Short, J. M., and Keller, M. (2002) Cultivating the uncultured. *Proc. Natl. Acad. Sci. U. S. A. 99*, 15681–15686.

(40) McClerklin, S. A., Lee, S. G., Harper, C. P., Nwumeh, R., Jez, J. M., and Kunkel, B. N. (2018) Indole-3-acetaldehyde dehydrogenase-dependent auxin synthesis contributes to virulence of Pseudomonas syringae strain DC3000. *PLoS Pathog. 14*, 1–24.

(41) Hitchcock, D. S., Fan, H., Kim, J., Vetting, M., Hillerich, B., Seidel, R. D., Almo, S. C., Shoichet, B. K., Sali, A., and Raushel, F. M. (2013) Structure-guided discovery of new deaminase enzymes. *J. Am. Chem. Soc. 135*, 13927–13933.

(42) Dhoke, G. V., Ensari, Y., Davari, M. D., Ruff, A. J., Schwaneberg, U., and Bocola, M. (2016) What's My Substrate? Computational Function Assignment of Candida parapsilosis ADH5 by Genome Database Search, Virtual Screening, and QM/MM Calculations. *J. Chem. Inf. Model. 56*, 1313–1323.

(43) Hermann, J. C., Marti-Arbona, R., Fedorov, A. A., Fedorov, E., Almo, S. C., Shoichet, B. K., and Raushel, F. M. (2007) Structure-based activity prediction for an enzyme of unknown function. *Nature 448*, 775–779.

(44) Zhao, S., Kumar, R., Sakai, A., Vetting, M. W., Wood, B. M., Brown, S., Bonanno, J. B., Hillerich, B. S., Seidel, R. D., Babbitt, P. C., Almo, S. C., Sweedler, J. V, Gerlt, J. A., Cronan, J. E., and Jacobson, M. P. (2013) Discovery of new enzymes and metabolic pathways by using structure and genome context. *Nature 502*, 698–702.

(45) Calhoun, S., Korczynska, M., Wichelecki, D. J., Francisco, B. S., Zhao, S., Rodionov, D. A., Vetting, M. W., Al-obaidi, N. F., Lin, H., Meara, M. J. O., Scott, D. A., Morris, J. H., Russel, D., Almo, S. C., Osterman, A. L., Gerlt, J. A., Jacobson, M. P., Shoichet, B. K., and Sali, A. (2018) Prediction of enzymatic pathways by integrative pathway mapping 1–27.

(46) Bible, A. N., Fletcher, S. J., Pelletier, D. A., Schadt, C. W., Jawdy, S. S., Weston, D. J., Engle, N. L., Tschaplinski, T., Masyuko, R., Polisetti, S., Bohn, P. W., Coutinho, T. A., Doktycz, M. J., and Morrell-Falvey, J. L. (2016) A carotenoid-deficient mutant in Pantoea sp. YR343, a bacteria isolated from the Rhizosphere of Populus deltoides, is defective in root colonization. *Front. Microbiol. 7*, 1–15.

(47) Estenson, K., Hurst, G. B., Standaert, R. F., Bible, A. N., Garcia, D., Chourey, K., Doktycz, M. J., and Morrell-Falvey, J. L. (2018) Characterization of Indole-3-acetic Acid Biosynthesis and the Effects of This Phytohormone on the Proteome of the Plant-Associated Microbe Pantoea sp. YR343. *J. Proteome Res. 17*, 1361–1374.

(48) Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., Tremblay, J., Engelbrektson, A., Kunin, V., del Rio, T. G., Edgar, R. C., Eickhorst, T., Ley, R. E., Hugenholtz, P., Tringe, S. G., and Dangl, J. L. (2012)

Defining the core *Arabidopsis thaliana* root microbiome. *Nature 488*, 86–90.

(49) Utturkar, S. M., Cude, W. N., Robeson, M. S., Yang, Z. K., Klingeman, D. M., Land, M. L., Allman, S. L., Lu, T.-Y. S., Brown, S. D., Schadt, C. W., Podar, M., Doktycz, M. J., and Pelletier, D. A. (2016) Enrichment of root endophytic bacteria from Populus deltoides and single-cell genomics analysis. *Appl. Environ. Microbiol. 82*, AEM.01285-16.

(50) Utturkar, S. M., Cude, W. N., Robeson, M. S., Yang, Z. K., Klingeman, D. M., and Land, M. L. (2016) Enrichment of Root Endophytic Bacteria from Populus deltoides and *82*, 5698–5708.

(51) Spaepen, S., Vanderleyden, J., and Remans, R. (2007) Indole-3-acetic acid in microbial and microorganism-plant signaling. *FEMS Microbiol. Rev. 31*, 425–448.

(52) Kelly, L. A., Mezulis, S., Yates, C., Wass, M., and Sternberg, M. (2015) The Phyre2 web portal for protein modelling, prediction, and analysis. *Nat. Protoc. 10*, 845–858.

(53) Wass, M. N., Kelley, L. A., and Sternberg, M. J. E. (2010) 3DLigandSite : predicting ligand-binding sites using similar structures *38*, 469–473.

(54) Deng, N., Forli, S., He, P., Perryman, A., Wickstrom, L., Vijayan, R. S. K., Tiefenbrunn, T., Stout, D., Gallicchio, E., Olson, A. J., and Levy, R. M. (2015) Distinguishing binders from false positives by free energy calculations: Fragment screening against the flap site of HIV protease. *J. Phys. Chem. B 119*, 976–988.

(55) Fujimoto, M. S., Suvorov, A., Jensen, N. O., Clement, M. J., and Bybee, S. M. (2016) Detecting false positive sequence homology: A machine learning approach. *BMC Bioinformatics 17*, 1–11.

(56) Houston, D. R., and Walkinshaw, M. D. (2013) Consensus docking: Improving the reliability of docking in a virtual screening context. *J. Chem. Inf. Model. 53*, 384–390.

(57) Stigliani, J. L., Bernardes-Génisson, V., Bernadou, J., and Pratviel, G. (2012) Cross-docking study on InhA inhibitors: A combination of Autodock Vina and PM6-DH2 simulations to retrieve bio-active conformations. *Org. Biomol. Chem. 10*, 6341–6349.

(58) Wang, Z., Sun, H., Yao, X., Li, D., Xu, L., Li, Y., Tian, S., and Hou, T. (2016) Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: The prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys. 18*, 12964–12975.

(59) Keiser, M. J., Roth, B. L., Armbruster, B. N., Ernsberger, P., Irwin, J. J., and Shoichet, B. K. (2007) Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol. 25*, 197–206.

(60) Durrant, J., and McCammon, J. A. (2011) Molecular dynamics simulations and drug discovery. *BMC Biol. 9*, 1–9.

(61) Spyrakis, F., Benedetti, P., Decherchi, S., Rocchia, W., Cavalli, A., Alcaro, S., Ortuso, F., Baroni, M., and Cruciani, G. (2015) A Pipeline to Enhance Ligand Virtual Screening: Integrating Molecular Dynamics and Fingerprints for Ligand and Proteins. *J. Chem. Inf. Model. 55*, 2256–2274.

(62) Stewart, E. J. (2012) Growing unculturable bacteria. *J. Bacteriol. 194*, 4151–4160.

(63) Lagier, J. C., Edouard, S., Pagnier, I., Mediannikov, O., Drancourt, M., and Raoult, D. (2015) Current and past strategies for bacterial culture in clinical microbiology. *Clin. Microbiol. Rev. 28*, 208–236.

(64) Jeon, W. B., Aceti, D. J., Bingman, C. A., Vojtik, F. C., Olson, A. C., Ellefson, J. M., McCombs, J. E., Sreenath, H. K., Blommel, P. G., Seder, K. D., Burns, B. T., Geetha, H. V., Harms, A. C., Sabat, G., Sussman, M. R., Fox, B. G., and Phillips, G. N. (2005) High-throughput purification and quality assurance of Arabidopsis thaliana proteins for eukaryotic structural genomics. *J. Struct. Funct. Genomics 6*, 143–147.

(65) Lesley, S. A. (2001) High-throughput proteomics: Protein expression and purification in the postgenomic world. *Protein Expr. Purif. 22*, 159–164.

(66) Kim, Y., Babnigg, G., Jedrzejczak, R., Eschenfeldt, W. H., Li, H., Maltseva, N., Hatzos-Skintges, C., Gu, M., Makowska-Grzyska, M., Wu, R., An, H., Chhor, G., and Joachimiak, A. (2011) High-throughput protein purification and quality assessment for crystallization. *Methods 55*, 12–28.

(67) Lal, P. B., Schneider, B. L., Vu, K., and Reitzer, L. (2014) The redundant aminotransferases in lysine and arginine synthesis and the extent of aminotransferase redundancy in *E scherichia coli*. *Mol. Microbiol. 94*, 843–856.

(68) Notebaart, R. A., Szappanos, B., Kintses, B., Pal, F., Gyorkei, A., Bogos, B., Lazar, V., Spohn, R., Csorg , B., Wagner, A., Ruppin, E., Pal, C., and Papp, B. (2014) Network-level architecture and the evolutionary potential of underground metabolism. *Proc. Natl. Acad. Sci. 111*, 11762–11767.

(69) Altschul, S. F., Gish, W., Miller, W., Myers, E. E. W. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.*

(70) Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc. 10*, 845–858.

(71) Trott, O., and Olson, A. J. (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem. 31*, NA-NA.

(72) Castro-Guerrero, J., Romero, A., Aguilar, J. J., Xiqui, M. L., Sandoval, J. O., and Baca, B. E. (2012) The hisC1 gene, encoding aromatic amino acid aminotransferase-1 in Azospirillum brasilense Sp7, expressed in wheat. *Plant Soil 356*, 139–150.

(73) Duca, D., Lorv, J., Patten, C. L., Rose, D., and Glick, B. R. (2014) Indole-3-acetic acid in plant–microbe interactions. *Antonie Van Leeuwenhoek 106*, 85–125.

(74) Ryu, R. J., and Patten, C. L. (2008) Aromatic amino acid-dependent expression of indole-3-pyruvate decarboxylase is regulated by tyrr in Enterobacter cloacae UW5. *J. Bacteriol. 190*, 7200–7208.

(75) Rivera, D., Revale, S., Molina, R., Gualpa, J., Puente, M., Maroniche, G., Paris, G., Baker, D., Clavijo, B., McLay, K., Spaepen, S., Perticari, A., Vazquez, M., Wisniewski-Dyé, F., Watkins, C., Martínez-Abarca, F., Vanderleyden, J., and Cassán, F. (2014) Complete Genome Sequence of the Model Rhizosphere Strain Azospirillum brasilense Az39, Successfully Applied in Agriculture. *Genome Announc. 2*, e00683-14.

(76) Yamada, T., Palm, C. J., Brooks, B., and Kosuge, T. (1985) Nucleotide sequences of the Pseudomonas savastanoi indoleacetic acid genes show homology with Agrobacterium tumefaciens T-DNA. *Proc. Natl. Acad. Sci. U. S. A. 82*, 6522–6526.

(77) Hull, A. K., Vij, R., and Celenza, J. L. (2000) Arabidopsis cytochrome P450s that catalyze the first step of tryptophan-dependent indole-3-acetic acid biosynthesis. *Proc. Natl. Acad. Sci. U. S. A. 97*, 2379–2384.

(78) Kato, Y., Yoshida, S., and Asano, Y. (2005) Polymerase chain reaction for identification of aldoxime dehydratase in aldoxime- or nitrile-degrading microorganisms. *FEMS Microbiol. Lett. 246*, 243–249.

(79) Kato, Y., Yoshida, S., Xie, S.-X., and Asano, Y. (2004) Aldoxime dehydratase co-existing with nitrile hydratase and amidase in the iron-type nitrile hydratase-producer Rhodococcus sp. N-771. *J. Biosci. Bioeng. 97*, 250–259.

(80) Williams, B. B., Van Benschoten, A. H., Cimermancic, P., Donia, M. S., Zimmermann, M., Taketani, M., Ishihara, A., Kashyap, P. C., Fraser, J. S., and Fischbach, M. A. (2014) Discovery and characterization of gut microbiota decarboxylases that can produce the neurotransmitter tryptamine. *Cell Host Microbe 16*, 495–503.

(81) Peñalber-Johnstone, C., Ge, X., Tran, K., Selock, N., Sardesai, N., Gurramkonda, C., Pilli, M., Tolosa, M., Tolosa, L., Kostov, Y., Frey, D. D., and Rao, G. (2017) Optimizing cell-free protein expression in CHO: Assessing small molecule mass transfer effects in various reactor configurations. *Biotechnol. Bioeng. 114*, 1478–1486.

(82) Bujara, M., Mperli, M., Pellaux, R., Heinemann, M., Panke, S., Schümperli, M., Pellaux, R., Heinemann, M., and Panke, S. (2011) Optimization of a blueprint for *in vitro* glycolysis by metabolic real-time analysis. *Nat. Chem. Biol. 7*, 271–277.

(83) Toya, Y., and Shimizu, H. (2013) Flux analysis and metabolomics for systematic metabolic engineering of microorganisms. *Biotechnol. Adv. 31*, 818–826.

(84) Rollin, J. A., Tam, T. K., and Zhang, Y. H. P. (2013) New biotechnology paradigm: Cell-free biosystems for biomanufacturing. *Green Chem. 15*, 1708–1719.

(85) Zhang, Y. H. P., Evans, B. R., Mielenz, J. R., Hopkins, R. C., and Adams, M. W. W. (2007) High-yield hydrogen production from starch and water by a synthetic enzymatic pathway. *PLoS One 2*, 2–7.

(86) Wang, H. H., Huang, P.-Y., Xu, G., Haas, W., Marblestone, A., Li, J., Gygi, S. P., Forster, A. C., Jewett, M. C., and Church, G. M. (2012) Multiplexed in Vivo His-Tagging of Enzyme Pathways for *in vitro* Single-Pot Multienzyme Catalysis. *ACS Synth. Biol. 1*, 43–52.

(87) Villarreal, F., Contreras-Llano, L. E., Chavez, M., Ding, Y., Fan, J., Pan, T., and Tan, C. (2018) Synthetic microbial consortia enable rapid assembly of pure translation machinery. *Nat. Chem. Biol. 14*, 29–35.

(88) Kim, D. M., and Swartz, J. R. (2001) Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnol. Bioeng. 74*, 309–316.

(89) Woodrow, K. A., and Swartz, J. R. (2007) A sequential expression system for high-throughput functional genomic analysis. *Proteomics 7*, 3870–3879.

(90) Zhang, X., Tervo, C. J., and Reed, J. L. (2016) Metabolic assessment of E. coli as a Biofactory for commercial products. *Metab. Eng. 35*, 64–74.

(91) Sun, Z. Z., Hayes, C. a, Shin, J., Caschera, F., Murray, R. M., and Noireaux, V. (2013) Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp.* e50762.

(92) Failmezger, J., Rauter, M., Nitschel, R., Kraml, M., and Siemann-Herzberg, M. (2017) Cell-free protein synthesis from non-growing, stressed Escherichia coli. *Sci. Rep. 7*, 16524.

(93) Nyström, T. (2004) Stationary-Phase Physiology. *Annu. Rev. Microbiol. 58*, 161–181.

(94) Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., and Heinemann, M. (2016) The quantitative and condition-dependent Escherichia coli proteome. *Nat. Biotechnol. 34*, 104–110.

(95) Washburn, M. P., Wolters, D., and Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol. 19*, 242–247.

(96) Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol. 428*, 726–731.

(97) Rohwer, J. M., Postma, P. W., Kholodenko, B. N., and Westerhoff, H. V. (1998) Implications of macromolecular crowding for signal transduction and

metabolite channeling. *Proc. Natl. Acad. Sci. 95*, 10547–10552.

(98) Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. 102*, 15545–15550.

(99) Cunningham, D. S., Liu, Z., Domagalski, N., Koepsel, R. R., Ataai, M. M., and Domach, M. M. (2009) Pyruvate Kinase-Deficient Escherichia coli Exhibits Increased Plasmid Copy Number and Cyclic AMP Levels. *J. Bacteriol. 191*, 3041–3049.

(100) Rolfe, M. D., Rice, C. J., Lucchini, S., Pin, C., Thompson, A., Cameron, A. D. S., Alston, M., Stringer, M. F., Betts, R. P., Baranyi, J., Peck, M. W., and Hinton, J. C. D. (2012) Lag Phase Is a Distinct Growth Phase That Prepares Bacteria for Exponential Growth and Involves Transient Metal Accumulation. *J. Bacteriol. 194*, 686–701.

(101) Noor, E., Eden, E., Milo, R., and Alon, U. (2010) Central Carbon Metabolism as a Minimal Biochemical Walk between Precursors for Biomass and Energy. *Mol. Cell 39*, 809–820.

(102) Goerke, A. R., and Swartz, J. R. (2008) Development of Cell-Free Protein Synthesis Platforms for Disulfide Bonded Proteins *99*, 351–367.

(103) Y., L., J., C., and S.-Y., L. (2001) Biotechnological production of pyruvic acid. *Appl. Microbiol. Biotechnol. 57*, 451–459.

(104) Maleki, N., and Eiteman, M. (2017) Recent Progress in the Microbial Production of Pyruvic Acid. *Fermentation 3*, 8.

(105) Zhang, Y., Wen, Z., Washburn, M. P., and Florens, L. (2010) Refinements to Label Free Proteome Quantitation: How to Deal with Peptides Shared by Multiple Proteins. *Anal. Chem. 82*, 2272–2281.

(106) Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K., Florens, L., and Washburn, M. P. (2006) Statistical Analysis of Membrane Proteome Expression Changes in Saccharomyces c erevisiae. *J. Proteome Res. 5*, 2339–2347.

(107) Guterl, J. K., Garbe, D., Carsten, J., Steffler, F., Sommer, B., Reiße, S., Philipp, A., Haack, M., Rühmann, B., Koltermann, A., Kettling, U., Brück, T., and Sieber, V. (2012) Cell-free metabolic engineering: Production of chemicals by minimized reaction cascades. *ChemSusChem 5*, 2165–2172.

(108) Shin, J., and Noireaux, V. (2012) An E. coli cell-free expression toolbox: Application to synthetic gene circuits and artificial cells. *ACS Synth. Biol. 1*, 29–41.

(109) Wang, H. H., Huang, P. Y., Xu, G., Haas, W., Marblestone, A., Li, J., Gygi, S. P., Forster, A. C., Jewett, M. C., and Church, G. M. (2012) Multiplexed in vivo his-tagging of enzyme pathways for *in vitro* single-pot multienzyme catalysis. *ACS Synth. Biol. 1*, 43–52.

(110) Rollin, J. A., Del Campo, J. M., Myung, S., Sun, F., You, C., Bakovic, A., Castro, R., Chandrayan, S. K., Wu, C. H., Adams, M. W. W., Senger, R. S., and Zhang, Y. H. P. (2015) High-yield hydrogen production from biomass by *in vitro* metabolic engineering: Mixed sugars coutilization and kinetic modeling. *Proc. Natl. Acad. Sci. U. S. A. 112*, 4964–4969.

(111) Wiegand, D. J., Lee, H. H., Ostrov, N., and Church, G. M. (2019) Cell-free Protein Expression Using the Rapidly Growing Bacterium Vibrio natriegens. *J. Vis. Exp. 176*, 139–148.

(112) Tinafar, A., Jaenes, K., and Pardee, K. (2019) Synthetic Biology Goes Cell-Free. *BMC Biol. 17*, 1–14.

(113) Lu, Y. (2017) Cell-free synthetic biology: Engineering in an open world. *Synth. Syst. Biotechnol. 2*, 23–27.

(114) Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., and Aebersold, R. (2016) The quantitative and condition-dependent Escherichia coli proteome *34*, 104–110.

(115) Dai, Z., and Nielsen, J. (2015) Advancing metabolic engineering through systems biology of industrial microorganisms. *Curr. Opin. Biotechnol. 36*, 8–15.

(116) T. Rydzak*, D. Garcia, M. Sladek, D.M. Klingeman, S. D. B. and A. G. (2016) Understanding nitrogen assimilation in Clostridium thermocellum for bioengineering. *2016 SIMB Annu. Meet. Exhib.*

(117) Lougheed, K. E. A., Bennett, M. H., and Williams, H. D. (2014) An in vivo crosslinking system for identifying mycobacterial protein-protein interactions. *J. Microbiol. Methods 105*, 67–71.

(118) Halliwell, C. M., Morgan, G., Ou, C. P., and Cass, A. E. G. (2001) Introduction of a (poly)histidine tag in L-lactate dehydrogenase produces a mixture of active and inactive molecules. *Anal. Biochem. 295*, 257–261.

(119) Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., and Church, G. M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature 460*, 894–8.

(120) Ronda, C., Pedersen, L. E., Sommer, M. O. A., and Nielsen, A. T. (2016) CRMAGE: CRISPR Optimized MAGE Recombineering. *Sci. Rep. 6*, 19452.

(121) Silva, F., Queiroz, J. A., and Domingues, F. C. (2012) Evaluating metabolic stress and plasmid stability in plasmid DNA production by Escherichia coli. *Biotechnol. Adv. 30*, 691–708.

(122) Wang, H. H., and Church, G. M. (2011) Multiplexed genome engineering and genotyping methods: Applications for synthetic biology and metabolic engineering. *Methods Enzymol.* 1st ed. Elsevier Inc.

(123) Booth, W. T., Schlachter, C. R., Pote, S., Ussin, N., Mank, N. J., Klapper, V., Offermann, L. R., Tang, C., Hurlburt, B. K., and Chruszcz, M. (2018) Impact of an N-terminal polyhistidine tag on protein thermal stability. *ACS Omega 3*, 760–768.

(124) Wingfield, P. T. (2015) Overview of the Purification of Recombinant Proteins. *Curr. Protoc. Protein Sci. 80*, 289–313.

(125) Stevenson, B. J., Liu, J. W., Kuchel, P. W., and Ollis, D. L. (2012) Fermentative glycolysis with purified Escherichia coli enzymes for *in vitro* ATP production and evaluating an engineered enzyme. *J. Biotechnol. 157*, 113–123.

(126) Doberenz, C., Zorn, M., Falke, D., Nannemann, D., Hunger, D., Beyer, L., Ihling, C. H., Meiler, J., Sinz, A., and Sawers, R. G. (2014) Pyruvate formate-lyase interacts directly with the formate channel FocA to regulate formate translocation. *J. Mol. Biol. 426*, 2827–2839.

(127) Birkenstock, T., Liebeke, M., Winstel, V., Krismer, B., Gekeler, C., Niemiec, M. J., Bisswanger, H., Lalk, M., and Peschel, A. (2012) Exometabolome analysis identifies pyruvate dehydrogenase as a target for the antibiotic triphenylbismuthdichloride in multiresistant bacterial pathogens. *J. Biol. Chem. 287*, 2887–2895.

(128) Rai, J., Pemmasani, J. K., Voronovsky, A., Jensen, I. S., Manavalan, A., Nyengaard, J. R., Golas, M. M., and Sander, B. (2014) Strep-tag II and Twin-Strep Based Cassettes for Protein Tagging by Homologous Recombination and Characterization of Endogenous Macromolecular Assemblies in Saccharomyces cerevisiae. *Mol. Biotechnol. 56*, 992–1003.

(129) Junttila, M. R., Saarinen, S., Schmidt, T., Kast, J., and Westermarck, J. (2005) Single-step Strep-tag® purification for the isolation and identification of protein complexes from mammalian cells. *Proteomics 5*, 1199–1203.

(130) Dalia, T. N., Hayes, C. A., Stolyar, S., Marx, C. J., McKinlay, J. B., and Dalia, A. B. (2017) Multiplex Genome Editing by Natural Transformation (MuGENT) for Synthetic Biology in Vibrio natriegens. *ACS Synth. Biol. 6*, 1650–1655.

(131) Mosberg, J. A., Lajoie, M. J., and Church, G. M. Lambda Red Recombineering in Escherichia coli Occurs Through a Fully Single-Stranded Intermediate.

(132) Gur, E., and Sauer, R. T. (2008) Evolution of the ssrA degradation tag in Mycoplasma: Specificity switch to a different protease. *Proc. Natl. Acad. Sci. 105*, 16113–16118.

(133) Cameron, D. E., and Collins, J. J. (2014) Tunable protein degradation in bacteria. *Nat. Biotechnol. 32*, 1276–1281.

(134) Nyerges, Á., Csörgő, B., Nagy, I., Bálint, B., Bihari, P., Lázár, V., Apjok, G., Umenhoffer, K., Bogos, B., Pósfai, G., and Pál, C. (2016) A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proc. Natl. Acad. Sci. 113*, 2502–2507.

(135) Zhou, L., Zuo, Z. R., Chen, X. Z., Niu, D. D., Tian, K. M., Prior, B. A., Shen, W., Shi, G. Y., Singh, S., and Wang, Z. X. (2011) Evaluation of genetic manipulation strategies on d-lactate production by Escherichia coli. *Curr.*

Microbiol. 62, 981–989.

(136) Richards, G. R., Patel, M. V., Lloyd, C. R., and Vanderpool, C. K. (2013) Depletion of glycolytic intermediates plays a key role in glucose-phosphate stress in escherichia coli. *J. Bacteriol. 195*, 4816–4825.

(137) Kabir, M. M., Pei, Y. H., and Shimizu, K. (2005) Effect of ldhA gene deletion on the metabolism of Escherichia coli based on gene expression, enzyme activities, intracellular metabolite concentrations, and metabolic flux distribution. *Biochem. Eng. J. 26*, 1–11.

(138) Schutte, K. M., Fisher, D. J., Burdick, M. D., Mehrad, B., Mathers, A. J., Mann, B. J., Nakamoto, R. K., and Hughes, M. A. (2015) Escherichia coli pyruvate dehydrogenase complex is an important component of CXCL10-mediated antimicrobial activity. *Infect. Immun. 84*, 320–328.

(139) Caschera, F., and Noireaux, V. (2014) Integration of biological parts toward the synthesis of a minimal cell. *Curr. Opin. Chem. Biol. 22*, 85–91.

(140) Garcia, D. C., Dinglasan, J. L. N., Shrestha, H., Abraham, P. E., Hettich, R. L., and Doktycz, M. J. (2020) A Lysate Proteome Engineering Strategy for Enhancing Cell-Free Metabolite Production. *bioRxiv* 2020.04.05.026393.

Vita

David C. Garcia was born in Winfield, Illinois to Abel Garcia and Reyna Garcia. He was raised in Michoacán, Mexico and West Chicago, Illinois. He attended Ripon College, graduating with a B.A. in Chemistry in 2014. He pursued his doctorate in Energy Science and Engineering at the University of Tennessee Knoxville and Oak Ridge National Laboratory. His doctoral research developing cell-free systems for natural product synthesis was overseen by Dr. Mitchel J. Doktycz.