

**Cell-free Metabolic Engineering Strategies for
Accelerated Biomanufacturing**

**A Dissertation Presented for the
Doctor of Philosophy
Degree**

The University of Tennessee, Knoxville

Jaime Lorenzo Naval Dinglasan

May 2023

Copyright © 2023 by Jaime Lorenzo Naval Dinglasan.

All rights reserved

DEDICATION

I dedicate this dissertation to my family.

To my parents, Jim and Cay, who are the best role models. Jim, you are the reason I am pursuing a career in science. Cay, my grit, determination, and, most importantly, balance come from you. I hope to be as accomplished as the both of you are.

To my siblings—Marco, Maika, Justine, and Elo—and to Jackson. Marco, I enjoy staying curious and know how to think creatively and critically because of you. Maika, you motivate me to strive for excellence in everything I do. Justine, you are the reason I value my education and hope to use it to help others. Elo, you have always inspired me to follow my own dreams. Jackson, thank you for teaching me to stay confident in my abilities.

To Lolo Cosme, your memory reminds me that a successful life is led with humility and compassion.

Mahal ko kayo.

ACKNOWLEDGEMENTS

I am extremely grateful for the opportunity to have performed this work with the facilities and scientists at Oak Ridge National Laboratory and the University of Tennessee. I am most thankful for the mentorship of Dr. Mitchel J. Doktycz, who has helped me grow tremendously as a scientist and as a person in the last four years. I would also like to thank my committee members—Dr. Jennifer Morrell-Falvey, Dr. Constance Bailey, and Dr. Paul Abraham—for also being my mentors and for being so involved in my research. I would like to acknowledge all of my collaborators as well for making this work possible. Finally, I am grateful for my friends and family in the Philippines and in America; you all are the best support system.

ABSTRACT

Biomanufacturing propels the bioeconomy. Accelerating bioeconomic growth thus requires the expedited development of biomanufacturing processes that can expand the current bioproduct portfolio. Lysate-based cell-free systems provide unique advantages for simplified metabolic pathway construction. Their central metabolic pathways and transcriptional-translational (TX-TL) machineries are free from genome regulation and are amenable to direct manipulation, enabling the streamlined construction of biomanufacturing processes. While their utility as prototyping platforms for accelerating cellular metabolic engineering has been demonstrated, the potential to rapidly build commercial “cell-free factories” capable of sophisticated bioconversion has not been fully realized. Lysates with high-yield pathways are projected to enable commercialized cell-free biomanufacturing of high-value chemicals. However, strategies that can be incorporated into frameworks for lysate pathway yield optimization rely on traditional cell-based metabolic engineering techniques that are cumbersome to extracts’ source strains (e.g., gene knockouts and cell-based overexpression) and thus lengthen design-build-test-learn (DBTL) cycles. The work in this dissertation introduces new strategies for cell-free pathway engineering that can benefit conversion yields in lysates, with a focus on 1) restructuring the endogenous metabolic proteome post-lysis and 2) optimizing the lysate TX-TL machinery for the cell-free overexpression of complex heterologous enzymes. By minimizing the involvement of cell-based engineering, these strategies enable faster endogenous and heterologous pathway build cycles compared to previous approaches. Part one describes the development of the first lysate flux-rewiring approach that enables conversion through a native pathway at 100% of the theoretical yield. Part two reports on the design and development of a plate reader assay for troubleshooting the expression of biosynthetic enzymes in lysates. The assay allows the generalizable screening of cell-free expression conditions with higher throughput than previous approaches and is leveraged to improve the lysate-based expression of natural product forming

megasyntases. Refining and integrating these approaches into cost-effective cell-free metabolic engineering (CFME) workflows will enable rapid high-yield metabolic pathway construction, advancing lysates as sustainable biomanufacturing platforms.

TABLE OF CONTENTS

1. Introduction	1
1.1. Biomanufacturing in a Growing Bioeconomy	2
1.2. Rapidly Expanding Biomanufacturing Capabilities with Cell-free Metabolic Engineering..	3
1.3. Efficient Tools for Engineering Efficient Lysate-based Biomanufacturing Platforms	6
1.4. Scope of Dissertation.....	10
1.5. Focusing Endogenous Metabolic Flux in Lysates Post-lysis	10
1.6. Improving Lysate-based CFE for Heterologous Pathway Building.....	11
1.7. Organization of Dissertation	12
2. Liquid chromatography coupled to refractive index or mass spectrometric detection for metabolite profiling in lysate-based cell-free systems.....	14
2.1. Abstract.....	15
2.2. Introduction	16
2.3. Methods	21
2.3.1. Starting, stopping, and processing time course CFME reactions for HPLC-RID quantification	21
2.3.2. Preparing the HPLC system for metabolite detection.....	23
2.3.3. Creating a method for the isocratic HPLC separation of organic fermentation products in OpenLab (Online).....	24
2.3.4. Creating a Sequence Table for autosampling and start the HPLC-RID system for data acquisition	25
2.3.5. Extracting and analyzing data post-run.....	25

2.3.6. Starting, stopping, and processing time course isotope tracing CFME reactions for LC-MS/MS quantification	26
2.3.7. Setting up the LC system for LC-MS/MS analysis	28
2.3.8. Creating a method on LC-MS/MS data acquisition and interpretation software for the LC system linked to Fourier Transform and Ion Trap Mass Spectrometers	28
2.3.9. Setting up a run sequence and starting the LC-MS/MS run	30
2.3.10. Consolidating files and searching for tentative annotations on MZmine 2.53	30
2.3.11. Calculating negative mode masses of ¹³ C-labeled glucose-derived metabolites and searching for the m/z features of these analytes in filtered data	32
2.4. Representative Results	33
2.4.1. HPLC-RID based analysis of lysate fermentative metabolism	33
2.4.2. Tracing carbon metabolism in lysates via LC-MS/MS	34
2.5. Discussion	40
3. A lysate proteome engineering strategy for enhancing cell-free metabolite production	49
3.1. Abstract	50
3.2. Introduction	51
3.3. Methods	54
3.3.1. Generation and validation of genome engineered strains using MAGE	54
3.3.2 Cell-free extract preparation protocol	55
3.3.3 Cell-free extract depletions	56
3.3.4 CFME reaction set-up	56
3.3.5 Proteomics sample preparation	57
3.3.6 LC-MS/MS analysis	57
3.3.7 Proteome database search	58

3.3.8 Proteomic data analysis	58
3.3.9 Metabolite measurements	59
3.4. Results and Discussion	59
3.4.1. Genome engineering enabled display of 6xHis-tags on endogenous pyruvate-consuming enzymes	59
3.4.2. Analysis of glucose metabolism and pyruvate production in enzyme-depleted extracts	64
3.4.3. Analysis of glucose metabolism and pyruvate production in enzyme-depleted extracts	66
3.4.4. Evaluation of modified lysate proteomes after selective enzyme depletion	70
3.4.5. Evaluating the feasibility of integrating StreptII®-tags into the genome of the model microbe <i>Vibrio natriegens</i>	74
3.5. Conclusions	75
4. Rewiring cell-free metabolic flux in e. Coli lysates using a block-push-pull approach	84
4.1. Abstract	85
4.2. Introduction	86
4.3. Methods	91
4.3.1. Cell Culture	91
4.3.2. Lysate Preparation	91
4.3.3. Lysate Depletion	92
4.3.4. CFME Reaction Preparation	92
4.3.5. Metabolite Analyses	93
4.4. Results and Discussion	94

4.4.1. Blocking flux-competing pathways via selective depletion of LdhA and PflB proteins in lysates redirects carbon flow to ethanol.	94
4.4.2. Optimized reaction and cultivation conditions push and pull flux through the insulated pathway and improve ethanol yield in blocked lysates	99
4.4.3. Maximum glucose-to-ethanol conversion is consistent in the optimized system	107
4.5. Conclusions	112
5. Applying a reporter system to investigate and optimize the lysate-based cell-free expression of nonribosomal peptide synthetases.....	123
5.1. Abstract.....	124
5.2. Introduction	125
5.3. Methods	128
5.3.1. Strains and plasmids	128
5.3.2. <i>In vivo</i> expression and purification of BpsA-TC	129
5.3.3. <i>In vivo</i> indigoidine production and purification	130
5.3.4. Cell-free extract preparation.....	131
5.3.5. CFE reaction preparation	131
5.3.6. Indigoidine quantitation in lysates with absorbance measurements	132
5.3.7. Comparative quantification of TC-based fluorescence measurements.....	133
5.3.8. BpsA concentration estimation using TC/FIAsh measurements.....	134
5.3.9. TycA and Pys concentration estimation	135
5.3.10. SDS-PAGE gel and Western Blot analyses.....	135
5.3.11. Quantification of active sfGFP	136
5.4. Results and Discussion	136
5.4.1. Cell-free expression of functional BpsA-TC in <i>E. coli</i> lysates.....	136

5.4.2. Optimization of a method for TC-based fluorescence measurement in an <i>E. coli</i> lysate CFE system.....	139
5.4.3. Unique reaction conditions are required for the full-length synthesis of different functional proteins.....	145
5.4.4. Nonribosomal peptide production improves with optimized NRPS expression	154
5.4.5. Conditions optimized for BpsA allow or improve the CFE of other single-module NRPSs.....	156
5.5. Conclusions	157
6. Concluding statements	174
References.....	181
Vita	208

LIST OF TABLES

Table 2.1. List of detected ¹³ C-labeled metabolites, retention times (aligned across samples using MZmine), theoretical fully ¹³ C-labeled negative mode m/z values, m/z values of detected features, and calculated mass errors.....	48
Table 3.1. Gene and protein information for MAGE targets with a potential effect on pyruvate metabolism.....	60
Table 3.2. Strains created for this study.....	63
Table 3.3. MAGE oligos use for this study. Phosphorothioated bases are noted with asterisks.	82
Table 3.4. MASC-PCR oligos used for this study.	82
Table 4.1. Number of cycles needed to wash ethanol off different resin suspension volumes with S30 buffer (for 200 µL lysate).	116
Table 5.1. Plasmid and gene sequences used in this study.....	168
Table 5.2. Primers used in this study; highlighted nucleotides are for the point mutation in the A domain: E315A.	173

LIST OF FIGURES

Figure 2.1. A putative metabolic model of lysates derived from <i>E. coli</i> BL21DE3-Star growing exponentially in high glucose concentrations.....	20
Figure 2.2. HPLC-RID data for glucose consumption and fermentative end-product synthesis in CFME reactions prepared with <i>E. coli</i> crude extracts.	35
Figure 2.3. Time course trends of ¹³ C ₆ -glucose and ¹³ C-labeled glycolytic intermediates in <i>E. coli</i> lysate CFME.	37
Figure 2.4. Time course trends of intermediates and end-products in ¹³ C ₆ -glucose fermentation in <i>E. coli</i> lysate CFME.....	38
Figure 2.5. Time course trends of detected ¹³ C ₆ -glucose derived amino acids in <i>E. coli</i> lysate CFME.	39
Figure 2.6. Time course trends of ¹³ C ₆ -glucose derived pentose phosphate pathway intermediates in <i>E. coli</i> lysate CFME.....	41
Figure 2.7. Representative HPLC-RID chromatogram showing peaks for major fermentative products in a CFME reaction incubated at 37 °C for 24 hrs.....	46
Figure 2.8. Representative mass spectra for ¹³ C-labeled metabolites in a CFME reaction incubated at 37 °C for 1 h.	47
Figure 3.1 Overview of approaches to preparing lysates for cell free metabolic engineering. ...	53
Figure 3.2. Glycolysis and engineered pathway nodes showing the location of the modified enzymes PflB, LdhA, PpsA, AceE.	60
Figure 3.3. Source strain multiplex genome engineering and expected metabolic phenotypes of derived lysates post-depletion.	62
Figure 3.4. Relative changes between nondepleted and depleted versions of the lysates.	65
Figure 3.5. Proteomic analysis of control cell-free extracts without depletion (blue), and a depleted cell-free extract (orange), and the elutions from the depleted extracts (gray).....	71

Figure 3.6. Growth rate and terminal OD600 as measured using a WPA CO 8000 Cell Density Meter (Biochrom Ltd., Cambridge, UK) for each of the four main strains generated in this study.	76
Figure 3.7. Glucose consumption and fermentation product concentrations over time in modified lysates.	77
Figure 3.8. Comparison of pyruvate consumption using depleted WT and $\Delta 4$ extracts.	78
Figure 3.9. Anti-6xHis western blot showing successful pull-downs of target proteins.	79
Figure 3.10. StreptII®-tagged glycolytic enzymes in modified <i>V. natriegens</i> strain, ST11.	80
Figure 3.11. Growth curves and metabolite data of strain ST11.	81
Figure 4.1. Target pathway and workflow for flux optimization.	90
Figure 4.2. Optimization of the post-lysis lysate depletion method to selectively remove 6xHis-tagged PflB and LdhA and redirect lysate metabolic flux to ethanol.	96
Figure 4.3. Optimization of cofactor concentrations in reactions prepared with blocked lysate to drive carbon flux towards ethanol.	100
Figure 4.4. Optimization of reaction vessel and initial reaction pH to drive carbon flux towards ethanol in blocked lysates.	102
Figure 4.5. Optimization of source strain cultivation conditions to drive carbon flux towards ethanol in resulting lysate proteomes.	105
Figure 4.6. Summary of the additive improvements to % ethanol yield in the system due to conditions that optimally block, push, and pull flux.	108
Figure 4.7. Evaluation of maximal glucose to ethanol conversion efficiency in the optimized system.	110
Figure 4.8. Residual ethanol from resin suspension storage buffer carried over to time zero lysate reaction set-ups.	116
Figure 4.9. Contribution of PflB depletion to the improvement of ethanol yield in LdhA depleted lysates.	117

Figure 4.10. Additional NAD ⁺ was supplemented to reactions prepared with blocked lysate, 100 mM Glucose, 0 mM CoA and ATP, but had no improvement or significant effect on % ethanol yield from glucose beyond 2 mM NAD ⁺	118
Figure 4.11. Acetate production or consumption in reactions.....	119
Figure 4.12. Acetate production or consumption in the optimized system and its non-depleted lysate counterpart.	120
Figure 5.1. Indigoidine production via BpsA in CFE reactions.	137
Figure 5.2. Optimization of TC/FIAsH complex measurements in <i>E. coli</i> lysate-based CFE reactions.....	141
Figure 5.3. TC/FIAsH analysis of energy substrate requirements in NRPS CFE using BpsA as a model.....	147
Figure 5.4. TC/FIAsH analysis of magnesium requirements in NRPS CFE using BpsA as a model.	150
Figure 5.5. TC/FIAsH analysis of amino acid (AA) requirements in NRPS CFE using BpsA as a model.....	152
Figure 5.6. Conditions optimized for BpsA improve indigoidine production and the expression of similar NRPSs in lysates.....	155
Figure 5.7. Standard curves of indigoidine from purified indigoidine in BAP1 lysate mock reactions.....	160
Figure 5.8. SDS-PAGE gel of purified BpsA-TC expressed heterologously in <i>E. coli</i> BAP1. ...	161
Figure 5.9. Optimization of FIAsH concentration for TC/FIAsH measurement and validation in <i>E. coli</i> lysate-based reactions.	161
Figure 5.10. BpsA-TC CFE in different temperatures and <i>E. coli</i> lysates.	162
Figure 5.11. TC/FIAsH measurements in different reaction backgrounds.....	163
Figure 5.12. Sequence alignment of BpsA indicating the conserved glutamine involved in the ATP binding motif.....	164

Figure 5.13. Effects of A domain activity on BpsA's PEP usage and expression.	165
Figure 5.14. SDS-PAGE validation of magnesium concentration effects on BpsA expression.	166
Figure 5.15. SDS-PAGE validation of amino acid concentration effects on BpsA expression.	166
Figure 5.16. SDS-PAGE gel of elutions from the StreptII-tag purification of TycA.....	167

CHAPTER 1

INTRODUCTION

1.1. Biomanufacturing in a Growing Bioeconomy

Biotechnology is revolutionizing various industries (e.g., healthcare, agriculture, plastics, food, and energy), and is emerging in others (e.g., textiles and aquaculture)¹. As a result, the bioeconomy—which encompasses these economic sectors that are impacted by biologically derived products, services, and processes—is growing and positively influencing overall economic development^{2,3}. The United States bioeconomy accounted for about 5% of gross domestic product in 2016, and is projected to provide more than 1 million job opportunities by 2030^{3,4}. Globally, biologically derived products, services, and processes are estimated to have a direct economic impact of \$4 trillion annually over the next 10-20 years⁵. Further, the development of the bioeconomy is agreed to be sustainable since biological resources are renewable and can therefore reduce the global reliance on oil and gas^{6,7}. Biotechnological advances are thus positioned to enable a future global “circular bioeconomy” in which unsustainable harvesting practices (e.g., mining) are eliminated, and renewable resources serve as inputs in closed material resource loops^{8,9}. Given the benefits of the bioeconomy, nations are issuing strategies to support its growth like the United States’ National Biotechnology and Biomanufacturing Initiative². At the core of such strategies is the accelerated development of industrial biomanufacturing platforms^{2,4,10}.

Biomanufacturing is defined as the synthesis of commercially relevant molecules in biological systems such as cells and enzymes⁶. Historically, microbes have been commonly considered suitable for bioproduction because they are easy to cultivate and can metabolically convert inexpensive substrates (e.g., gases and sugars) to higher value chemicals and materials^{6,11,12}. Although the use of mixed microbial processes to carry out such conversions dates back thousands of years, the application of dedicated mono-cultures towards commercialized production only began in the early 1900s¹⁰. This was then rapidly scaled up through the mid 1900s due to high demands for certain metabolites (e.g., citric acid, penicillin) during war times¹⁰. To

help realize biomanufacturing's feasibility as a replacement to industrial petrochemical based processes, the late 1900s saw the emergence of the field of Metabolic Engineering¹⁰. Leveraging recombinant DNA technology and cell culture techniques, the goal of Metabolic Engineering is to transform a microbe into a "cell factory" that can carry out targeted conversions^{13,14}. Contingent to a cell factory's adoption for commercial manufacturing is its ability to form a desired molecule at a titer (amount of product per unit volume), productivity (rate of product formation per unit volume), and yield (amount of product formed per amount of substrate consumed) that is industrially relevant^{13,15}. Today, a few producer strains have been engineered with high productivities, titers, and yields, and are hence being applied towards the commercial-scale production of certain fine chemicals (e.g., pharmaceuticals and fragrances)¹⁶ and commodity chemicals (e.g., biofuels)⁷.

1.2. Rapidly Expanding Biomanufacturing Capabilities with Cell-free Metabolic Engineering

In principle, a much larger variety of chemicals and materials can be made by bioindustries. According to the McKinsey Global Institute, up to 60% of the global economy's factors of production could be biologically derived, the majority of which (e.g., bio-plastics) will have to be biomanufactured⁵. However, the deployment of producer strains that will broaden the range of industrial bioproducts is delayed by limitations that are inherent to engineering living systems¹⁷⁻²⁰. That is, rationally developing a high-performing microbial host that sustains target metabolite production while remaining viable requires numerous design, build, test, and learn (DBTL) cycles^{21,22}. Iterations of the build step, where a pathway design is constructed into a chassis, are particularly rate-limiting in rational metabolic engineering^{23,24}. This typically involves 1) rewiring endogenous metabolism via knockouts or gene regulation to redirect flux, and 2) expressing heterologous enzymes to overcome bottlenecks or to build in heterologous

pathways¹³. Such unnatural tasks are likely to result in metabolic burden or product cytotoxicity, among other phenomena, that can be deleterious to cell viability and performance²⁵. Constructing a producer strain with insignificantly compromised health phenotypes thus requires laborious strain optimization to balance cell growth with biomanufacturing objectives²⁵. This is a lengthy process even in model microbes, whose current engineering toolkits have been developed over decades²⁶. As a result, common workflows for building testable pathway designs in cells can take weeks in industrial workhorses like *E. coli* and several months in non-model organisms, constraining our biomanufacturing capabilities^{26,27}. On the other hand, next-generation platforms that enable shorter build stages, such as cell-free systems, can expedite the diversification and expansion of the commercial bio-product portfolio^{17,19–21,28–30}.

Cell-free or *in vitro* systems have the components necessary to carry out metabolism or protein expression outside of a living cell, or in a “cell-free” space³⁰. These are commonly generated either as “purified” systems by the *in vitro* reconstitution of purified parts (e.g., metabolic enzymes or the transcriptional-translational (TX-TL) machinery), or as crude cell extracts via cell lysis procedures that preserve the metabolic and TX-TL functionalities of the cytosolic proteome³⁰. In either case, the cell-free system serves as an open environment of biochemical reactions that are amenable to direct manipulation^{20,21,30}. Unlike cells, cell-free reactions can be altered without needing to transport substrates and products across an enclosing membrane barrier^{20,21,30}. Further, the active processes in these systems are not genome-regulated and are thus less dynamic, offering higher programming predictability^{20,21,30}. These advantages of cell-free platforms were first leveraged to uncover the fundamentals of the central dogma of molecular biology³¹. However, the growing realization that cell-free systems have the potential to also support applied research ventures backed recent efforts to promote their widespread and large-scale use^{30,32}. Among these applications is cell-free metabolic engineering (CFME), where cell-free systems are used for the construction of pathway designs that are time-consuming or impossible to build into microbial cell factories^{11,17,18,28,29}.

Cell-free systems provide unique engineering capabilities for simplifying metabolic pathway building, thereby permitting expedited DBTL cycles^{10,20,21,27}. Their lack of a cell membrane barrier allows the direct addition of metabolic enzymes, substrates, and cofactors into the open reaction environment for easy pathway construction and manipulation³⁰. Enzymes can also be expressed directly from a linear or plasmid DNA template that is dropped into cell-free systems with activated TX-TL machineries³⁰. After supplementation with DNA, cell-free expression (CFE) occurs in hours whereas transforming and selecting strains for *in vivo* enzyme synthesis can take days to weeks depending on the host chassis^{33–35}. The detachment of cell-free systems from genome-regulated survival functions (i.e., growth and maintenance) further simplifies pathway construction²⁷. Engineering cells towards target production objectives can be unpredictable and time-consuming due to complications on host health like metabolic burden and product cytotoxicity^{25,34,36–38}. On the other hand, cell-free systems do not allocate resources targeted for biosynthesis towards growth and maintenance, and they can overcome cellular toxicity limits^{20,30,39}. These features accelerate building stages in cell-free systems compared to their cell-based counterparts²². For example, commercially relevant titers of cytotoxic natural products like limonene, pinene, and alpha-sabinene have been achieved in an *in vitro* system (> 11 g/L), but the synthesis of these compounds through similar pathways in the best performing cell factory to date is still encumbered by cytotoxicity (< 3 g/L)⁴⁰. By streamlining metabolic engineering efforts, cell-free systems offer a potential route to the industrial synthesis of bio-based chemicals and materials that have not been produced efficiently in cells.

Cell-free systems have the added advantage of being freeze dried then stockpiled and transported at room temperature⁴¹. Upon rehydration, their bioprocesses can be reactivated for distributed manufacturing, which addresses limitations inherent to centralized production like the costs of maintaining and customizing large manufacturing plants as well as cold chain distribution⁴². In addition, distributed manufacturing enables synthesis at the point-of-care, which is especially important during times of supply chain disruption (e.g., global pandemics).

Lyophilized cell-free systems have already been leveraged to demonstrate the “just-add-water” production of therapeutic proteins^{41–43}. Engineering cell-free systems with sophisticated bioconversion pathways can enable the cheap, decentralized production of valuable small molecules from cheap feedstocks like glucose in the future⁴⁴.

1.3. Efficient Tools for Engineering Efficient Lysate-based Biomanufacturing Platforms

Between the two types of cell-free systems, lysates are better positioned to be next-generation platforms for converting low-cost feedstocks to high-value chemicals¹¹. Reconstituted enzyme systems suffer from cost-prohibitory enzyme purification and typically low cofactor turnover, which limit their scalability^{11,20,45}. Generating application-specific reconstituted systems with all the necessary metabolic modules is thus unlikely to be practical for commercial-scale production. In contrast, crude cell extracts are easy to prepare and already possess active central metabolic pathways, including cofactor regenerating reactions^{46,47}. Lysates can thus convert readily available substrates like glucose into valuable native metabolites, such as the pyruvate and acetyl-CoA precursors of central metabolism^{46,48–50}. Heterologous pathways that interface with these native metabolites can then be built into lysate-based cell-free systems to support the synthesis of commercially relevant end-products (e.g., 2,3-butanediol and styrene)^{46,51}. This has been done in different ways which include directly supplementing lysate reactions with purified enzymes or deriving enzyme-enriched extracts from heterologous expression hosts^{19,46,52,53}. Alternatively, the lysate TX-TL machinery itself can be activated for heterologous pathway expression to bypass laborious *in vivo* expression^{18,27}. Although cell-free metabolic yields from pathways constructed via cell-based expression rather than CFE can be higher⁵⁴, the latter is an attractive strategy for prototyping several pathway designs or when dealing with heterologous enzymes that are difficult to synthesize *in vivo*^{18,27}. Additionally, this process can be scaled up, as

evidenced by findings of similar lysate-based CFE expression profiles in 250 μ L and 100 L reactors⁵⁵.

Cell-free synthetic biology's role in revolutionizing biomanufacturing can be realized by cell extracts. However, their industrial implementation is still limited. Lysate-based cell-free systems are being applied to interrogate pathway designs which has proven useful for significantly shortening DBTL cycles from months-years to days-weeks in a non-model microbial cell factory²⁶. However, they are not yet economically viable for carrying out sophisticated bioconversions due to current cell processing and reagent costs¹¹. To address this, potential metabolic engineering solutions have been proposed¹¹. These include extending the shelf-lives of these systems' enzymes^{20,40}, building in more efficient cofactor regenerating modules^{56,57}, or engineering extracts to utilize free feedstocks (e.g. CO₂)¹¹ and cheaper alternative cofactors (e.g., non-canonical nicotinamides)^{58,59}. Based on calculations by Rasor et al., such innovations can advance lysate-based systems as lucrative platforms for high-value specialty chemical production, provided that these systems have maximum product yields¹¹.

Yield is an important metric in metabolic engineering and measures the efficiency of the system for converting a substrate into a product^{11,60}. Specifically, it is the amount of product formed per amount of substrate consumed. Even with a significant reduction in preparation costs, a lysate-based biomanufacturing platform dedicated to high-value target synthesis must have a high yield to be lucrative¹¹. However, industrially relevant yields achieved through CFME are rarely reported in extracts, even if these systems are capable of high titers^{46,51,53} and productivities^{46,61}. Engineering both endogenous and heterologous lysate metabolism is necessary to ensure carbon flow directly from substrate to product, which will improve yields in systems made to produce high-value molecules⁶⁰.

The endogenous metabolism in lysates is complex, retaining much of the pathway crosstalk that exists in the metabolic networks of their source strains²⁰. Thus, flux from a substrate is commonly redirected towards unwanted byproducts. To maximize yield, flux can be focused

through a targeted endogenous pathway by shedding this complexity (i.e., eliminating flux-competing reactions)^{47,62}. Altering lysate endogenous pathways to eliminate by-product formation permits higher pooling of desirable central metabolic precursor/s that will feed into heterologous pathways^{47,53,60,62}. Previous efforts to increase flux through endogenous lysate metabolism involve knocking out genes from source strains^{47,53,62}. However, this approach is highly constrained by source strain viability, given that genes selected for deletion typically contribute to the organism's central metabolism. Knocking out these reactions from the source strain compromise growth and may be lethal, so that a limited number of gene deletions are allowed without significant strain optimization¹³. Thus, this technique still requires long DBTL cycles that are characteristic of cell-based metabolic engineering campaigns.

High-value bioconversions typically necessitate the construction of heterologous metabolic pathways that feed off pooled endogenous precursors for conversion to a target product¹⁰. Overexpressing heterologous enzymes in a metabolic pathway is thus necessary to drive flux towards the desired metabolite with high yield⁴⁶. The common approach for doing this in lysate-based systems is by overexpressing these proteins in the source strain, then deriving enzyme-enriched extracts for cell-free conversion^{46,51,53}. Unfortunately, overexpression can drain the cell's physiological resources to make proteins, imposing metabolic burden on the organism²⁵. The physicochemical properties of metabolic enzymes that are non-native to the cell can also introduce complications in the host. For example, these proteins and their metabolic products may be cytotoxic, or these enzymes may cause cofactor imbalances when interfacing with endogenous metabolism^{23,24}.

A major advantage of lysates is that they can support rapid pathway construction because they are free from cell-based biochemical (i.e., resource competition between metabolite and biomass production) and biophysical (i.e., cell membrane barrier) constraints. They have the potential to be easy-to-build, sustainable manufacturing platforms which, in turn, can accelerate the expansion of the commercial bioproduct space. However, current CFME approaches to improve

an industrially relevant metric in these systems rely on techniques that can introduce deleterious phenotypes in source strains (i.e., gene knockouts and overexpression). These strategies reintroduce steps into CFME workflows that are constrained by host viability, which is counterintuitive to realizing streamlined pathway construction in cell-free systems. Efforts to improve pathway yield in lysate-based biomanufacturing systems would thus benefit from strategies that do not alter source strain metabolism. That is, restructuring lysate endogenous metabolism and enriching lysates with heterologous enzymes should be done via totally cell-free approaches.

The removal of select enzymes from the lysate itself instead of the cell is an opportunity for rewiring lysate endogenous metabolism without altering the source strain's growth phenotype. If this can be performed successfully, the number of reactions that can be eliminated from the lysate proteome would only be limited by the efficiency of the removal strategy. Further, changes to metabolic proteome in lysates should be more predictable and tractable because, unlike in cells, making such modifications does not induce the genome-regulated optimization of competing objectives (e.g., biomass production). Thus, complementing such an approach with knowledge of lysate endogenous metabolism will simplify efforts to model high-yield pathway designs in lysates. However, a post-lysis metabolic enzyme removal strategy has not yet been reported and the feasibility of this approach for future CFME campaigns requires investigation and demonstration.

CFE is a promising approach for rapidly building heterologous pathways into lysate-based biomanufacturing platforms since it is completely decoupled from cell growth¹⁸. It can also produce functionally active enzymes that are typically difficult to express *in vivo* due to size, complexity, or solubility^{30,63}. Thus, leveraging CFE for pathway overexpression bypasses the laborious optimization of source strains to produce non-native enzymes that can be deleterious to their health, accelerating pathway build stages. However, the dearth of information on how most biosynthetic enzymes are transcribed and translated in the cell-free environment is a barrier to CFE's application as a CFME tool beyond pathway prototyping. Investigating and optimizing

the synthesis of metabolic enzymes in lysates will enable the use of CFE for optimizing flux through heterologous pathways. The efficiency of an enzyme's expression and catalytic activity in lysates depends on the availability of supplemental resources (i.e., energy substrates, salts, amino acids, etc.) that energize the TX-TL machinery and other reaction conditions (e.g., temperature)⁶³⁻⁶⁵. In theory, reaction condition sets for optimal CFE vary based on the target enzyme's physicochemical properties which include its size, folding, and biosynthetic activity^{63,66}. Achieving the sufficient expression of different non-native biosynthetic enzymes via CFE will thus require sampling different reaction parameters first. Unfortunately, this process is currently bottlenecked by slow or inaccessible detection methods (e.g., SDS-PAGE or mass spectrometry). Generalizable reporter tools for studying and improving biosynthetic enzyme expression in lysates will accelerate overexpression optimization via CFE, enabling the quicker and easier construction of efficient lysate-based biomanufacturing platforms.

1.4. Scope of Dissertation

This dissertation aims to expand the current CFME toolkit for building high-yield pathway designs in extracts. The focus is on providing strategies for engineering endogenous and heterologous lysate metabolism that do not rely on deleterious cell-based engineering techniques, avoiding global alterations to source strain metabolism that are deleterious to host health. The approaches can help advance lysates as easy-to-build, sustainable biomanufacturing technologies which, in turn, can accelerate the expansion of the commercial bioproduct space.

1.5. Focusing Endogenous Metabolic Flux in Lysates Post-lysis

Rapid pathway optimization in lysates can be realized by rewiring endogenous metabolic flux post-lysis. To design high-yield pathways, the movement of carbon from common feedstocks

through lysate endogenous metabolism in lysates should first be understood. In line with this, potential channels through which carbon derived from glucose, a readily available and inexpensive substrate, can flow in the model *Escherichia coli* lysate are mapped out (Chapter 2). This investigation leverages liquid chromatography coupled to a mass spectrometric detector and fully ^{13}C -labeled glucose to track possible metabolic endpoints of glucose-derived carbon. To manipulate this carbon flow, a lysate proteome engineering strategy is developed (Chapter 3). This is demonstrated in model *Escherichia coli* lysates that are fed glucose and made to accumulate the central metabolic precursor pyruvate. The approach takes advantage of a genome engineering method to insert affinity tags into the termini of pyruvate-consuming enzymes, enabling the selective reduction of these proteins' concentrations in processed extracts. With an understanding of lysate metabolism and a proteome engineering strategy in-hand, the full depletion of target enzymes in an *E. coli* lysate can be achieved, as demonstrated with ethanol, a tractable end-product of endogenous metabolism (Chapter 4). The selective enzyme depletion approach is optimized and coupled with other CFME approaches to focus and drive carbon flux from glucose to ethanol, achieving maximum (100%) conversion yield through this native pathway. This is the first report of an engineering method that can confer a lysate's low-yield pathway with industrially relevant performance. Uniquely, the approach does not introduce gene knockouts, knockdowns, or overexpression and consequently does not impose observable metabolic burden on the source strain. This therefore represents a strategy for high-yield endogenous pathway construction in lysates that is free from cellular viability constraints.

1.6. Improving Lysate-based CFE for Heterologous Pathway Building

Heterologous pathway building and optimization via CFE, as opposed to cell-based engineering approaches, will be useful for rapidly constructing lysate-based biomanufacturing platforms that efficiently synthesize high-value molecules from pooled endogenous precursors.

To enable CFE as a tool for overexpressing metabolic enzymes in lysates, rapid reaction condition testing is necessary. A reporter-based strategy for measuring full-length protein expression is developed in *E. coli* lysates (Chapter 5). A tetracysteine (TC) peptide tag with activatable fluorescence is conjugated to the C-terminus of a biosynthetic enzyme. A method for activating and measuring the fluorescence of this tag in the noisy lysate environment is then optimized, allowing quick detection of full-length biosynthetic enzymes produced in lysate-based CFE reactions. This technique is generalizable and can be used to quickly measure heterologous metabolic enzyme CFE under different reaction conditions. This knowledge can be leveraged to overexpress biosynthetic enzymes at higher concentrations in lysates and drive flux through heterologous reactions. The approach is demonstrated with the investigation and optimization of nonribosomal peptide synthetase (NRPS) CFE. NRPSs synthesize nonribosomal peptides, natural products with diverse commercially relevant biological properties (i.e., antimicrobial, anticancer, immunosuppressive, and insecticidal activity)⁶⁷. CFE is being increasingly considered for NRPS synthesis since their large (> 100 kDa) and complex structures make them difficult to express *in vivo*^{28,68}. CFE conditions optimized via this technique for blue pigment synthetase A (BpsA), a model nonribosomal peptide synthetase (NRPS)⁶⁹, improve the expression of other NRPSs as well. Further, improving BpsA expression in lysates increases the production of its pigment product from glutamine, a central metabolite that can be endogenously made in lysates.

1.7. Organization of Dissertation

Part I of this dissertation encompasses Chapters 2-4 which describe the development of an effective lysate proteome engineering approach for restructuring endogenous metabolism in *E. coli* extracts. Chapter 2 is adapted from the article “Liquid Chromatography Coupled to Refractive Index or Mass Spectrometric Detection for Metabolite Profiling in Lysate-based Cell-free Systems” which was published in the *Journal of Visualized Experiments* in 2021. Chapter 3

is adapted from the article “A Lysate Proteome Engineering Strategy for Enhancing Cell-Free Metabolite Production” originally published in *Metabolic Engineering Communications* in 2021. Chapter 4 is entitled “Rewiring Cell-free Metabolic Flux in *E. coli* Lysates using a Block-Push-Pull Approach” and is currently under review for publication as a research article in *Synthetic Biology*. Part II of this manuscript consists of Chapter 5 and details the optimization of a reporter-based tool that enables higher heterologous expression of functional biosynthetic enzymes. It is adapted from the article “Applying Reporters to Investigate and Optimize Lysate-Based Cell-Free Expression for Nonribosomal Peptide Production” which is currently under review for publication as a research article in *ACS Synthetic Biology*. The Appendix section at the end of each chapter includes the Supplementary Information associated with each article. The References of each chapter are collated at the end of this dissertation.

CHAPTER 2

**LIQUID CHROMATOGRAPHY COUPLED TO REFRACTIVE
INDEX OR MASS SPECTROMETRIC DETECTION FOR
METABOLITE PROFILING IN LYSATE-BASED CELL-FREE
SYSTEMS**

A version of this chapter was originally published by Jaime Lorenzo N. Dinglasan, David T. Reeves, Robert L. Hettich, and Mitchel J. Doktycz:

Dinglasan, J. L. N., Reeves, D. T., Hettich, R. L., Doktycz, M. J. (2021) Liquid Chromatography Coupled to Refractive Index or Mass Spectrometric Detection for Metabolite Profiling in Lysate-based Cell-free Systems. *Journal of Visualized Experiments* (175), e62852, doi:10.3791/62852

This chapter has been adapted from its published format to accommodate new Figure, Table, and Equation enumeration. The Supplementary Information associated with this work may be found in the Appendix section of this chapter. All references are located at the end of the document.

2.1. Abstract

Engineering cellular metabolism for targeted biosynthesis can require extensive design-build-test-learn (DBTL) cycles as the engineer works around the cell's survival requirements. Alternatively, carrying out DBTL cycles in cell-free environments can accelerate this process and alleviate concerns with host compatibility. A promising approach to cell-free metabolic engineering (CFME) leverages metabolically active crude cell extracts as platforms for biomanufacturing and for rapidly discovering and prototyping modified proteins and pathways. Realizing these capabilities and optimizing CFME performance requires methods to characterize the metabolome of lysate-based cell-free platforms. That is, analytical tools are necessary for monitoring improvements in targeted metabolite conversions and in elucidating alterations to metabolite flux when manipulating lysate metabolism. Here, metabolite analyses using high-performance liquid chromatography (HPLC) coupled to either optical or mass spectrometric detection were applied

for characterizing metabolite production and flux in *E. coli* S30 lysates. Specifically, this report describes the preparation of samples from CFME lysates for HPLC analyses using refractive index detection (RID) to quantify the generation of central metabolic intermediates and by-products in the conversion of low-cost substrates (i.e., glucose) to various high-value products. The analysis of metabolite conversion in CFME reactions fed with ^{13}C -labeled glucose through reversed-phase liquid chromatography coupled to tandem mass spectrometry (MS/MS), a powerful tool for characterizing specific metabolite yields and lysate metabolic flux from starting materials, is also presented. Altogether, applying these analytical methods to CFME lysate metabolism enables the advancement of these systems as alternative platforms for executing faster or novel metabolic engineering tasks.

2.2. Introduction

Limitations in engineering microbes for chemical production can be addressed by recapitulating biochemical reactions *in vitro* where competing cellular survival functions are absent²⁰. Moreover, the open reaction environment (i.e., absence of a cell membrane) is more amenable to manipulation and is easier to monitor compared to live cells. This foundational concept of cell-free metabolic engineering (CFME) has been elegantly demonstrated by the reconstitution of metabolic pathways to synthesize valuable chemicals like hydrogen and monoterpenes with production metrics that are orders of magnitude higher than presented in microbial cell factories thus far^{20,40,70}. Methods for purifying whole pathways, however, are currently constrained by time and cost. Alternatively, cell-free metabolic systems can be derived from crude cell extracts through rapid and inexpensive methods relative to whole pathway reconstitution⁷¹. The central metabolism that is retained in cell extracts can be supplemented with energy substrates (e.g., glucose and enzymatic cofactors) and salts in buffered solutions to generate central metabolic precursors for over 24 h^{46,72}. Adding exogenous enzymes to the lysate-

based CFME reaction allows more complex bio-transformations of glucose into more valuable chemicals at high titers^{46,71,73}. Although yield tends to be compromised in these systems due to their cell-like metabolic complexity, unique methods to curate lysate proteomes for higher yield conversion have been and are being developed^{72,74}.

The ease of carrying out metabolic transformations in lysate-based cell-free systems makes these excellent platforms for either moving chemical manufacturing outside of the cell altogether or for prototyping new pathways with high throughput before building and testing these designs *in vivo*^{20,22}. For either application, tools for monitoring metabolic conversions or observing overall alterations to metabolic flux in lysates are integral to the advancement of CFME. High-performance liquid chromatography (HPLC) can be used to separate the chemical constituents of CFME reactions with high-resolution and can be coupled to optical or mass-spectrometric detectors for metabolite quantification^{48,72}. The underlying principle of HPLC is that analytes dissolved in a solvent (i.e., mobile phase) and pumped through a column will interact with the specific column packing material (i.e., stationary phase)⁷⁵. Depending on their chemical properties, these analytes exhibit varying retention times before they are eventually eluted from the stationary phase and carried by the mobile phase to a detector. This report details the preparation and analysis of *E. coli* lysate-based CFME reactions through HPLC-based methods leveraging RID and MS/MS detection.

HPLC coupled to refractive index detection (HPLC-RID) is a generally accessible method for quickly identifying central metabolic precursors and end-products. Briefly, RID measures how analytes change the deflection of light by the mobile phase⁷⁶. RID signals corresponding to target analytes in samples can then be quantified by comparisons with RID signals of standard solutions. In CFME applications, this mode of detection has been most commonly used with HPLC columns that separate compounds based on a combination of size exclusion and ligand exchange mechanisms, or ion-moderated partition chromatography^{46,72,74,77}. This particular technique is used to quickly quantify the consumption of sugar substrates like glucose as well as the formation

of fermentation products like succinate, lactate, formate, acetate, and ethanol in lysate-based CFME reactions⁷⁴. Recording the concentration changes of these compounds via HPLC has been useful for both elucidating the potential of crude cell extracts to pool central metabolic precursors and understanding how pathway flux is redirected through fermentative pathways during complex metabolic conversions from glucose in lysates^{46,74,78}. Seminal CFME studies in *E. coli* cell extracts confirm that fermentation compounds accumulate as end-products of glucose catabolism and also occur as unwanted by-products in lysates that overexpress exogenous enzymes^{46,47}. It is suggested that fermentative metabolism plays a necessary role in regenerating redox equivalents of cofactors (i.e., NAD(P)H and ATP) to sustain glycolytic reactions⁷⁴. Hence, an HPLC-based optical detection method designed to separate fermentation products is a useful and commonly applied tool when executing various lysate-based CFME tasks.

CFME can be implemented to accumulate metabolic end-products that are not carbohydrates, organic acids, or alcohols⁷¹. The measurement of intermediates that are consumed as quickly as they are synthesized may also be desirable⁴⁸. While HPLC-RID is accessible in terms of cost and difficulty, this method is constrained by its ability to only distinguish metabolites based on retention time. A broader range of metabolites can be analyzed when liquid chromatography is coupled to MS/MS detection (LC-MS/MS)⁷⁹. By this method, analytes in the mobile phase are ionized and differentially detected based on each molecule's mass and charge properties. Knowledge of both the metabolite's mass-to-charge (m/z) ratio and retention time on the column thus facilitates the separation of most metabolic intermediates and end-products with high resolution⁷⁹. This detection technique can also be coupled to nano-liquid chromatography, which affords much lower flow rates and sample injection volumes, allowing more sensitive detection of small molecules in the complex lysate background⁸⁰. LC-MS/MS can additionally be applied with isotope labeling since incorporated labels impart changes in analytes' m/z values⁸¹. Timepoint measurements extracted from a CFME reaction supplemented with a $^{13}\text{C}_6$ -glucose substrate can thus determine the end- or by-products derived specifically from supplemented

glucose. Although this isotope tracing method is not yet commonly applied in CFME studies, it is a powerful tool for understanding metabolic conversions in lysate-based CFME systems, specifically since salt counterions (i.e., acetate and glutamate) in these reactions are also catabolized as secondary substrates⁴⁹. Leveraging this technique can thus draw a comprehensive picture of glucose metabolism in lysates, which to this day is not completely understood. Here, the protocol details a method for nano-liquid chromatography coupled to nanoelectrospray ionization (nano ESI) MS/MS that can be used to interrogate a possible model of glucose metabolism, specifically in *E. coli* lysates (**Figure 2.1**). The model is based on reports of fermentative pathways and the pentose phosphate pathway being active in *E. coli* lysates derived from strains grown in rich media^{46,72,74,78}. The technique is additionally used to investigate amino acid production since current knowledge on amino acid anabolism from glucose in lysates is limited to a few examples such as the synthesis of aromatic amino acids⁷³. Given the mostly polar nature of end-products and intermediates in these pathways (i.e., organic acids, sugar phosphates, and amino acids), reverse-phased liquid chromatography was utilized here. This technique separates polar compounds by elution from a nonpolar stationary phase. These compounds were then ionized by nano ESI in negative ion mode which permits the detection of analytes with at least one negative elementary charge and is thus useful for detecting acidic compounds. This technique is employed here to analyze glucose-derived ¹³C-incorporating metabolites and demonstrates the utility of LC-MS/MS for understanding glucose metabolism in lysates.

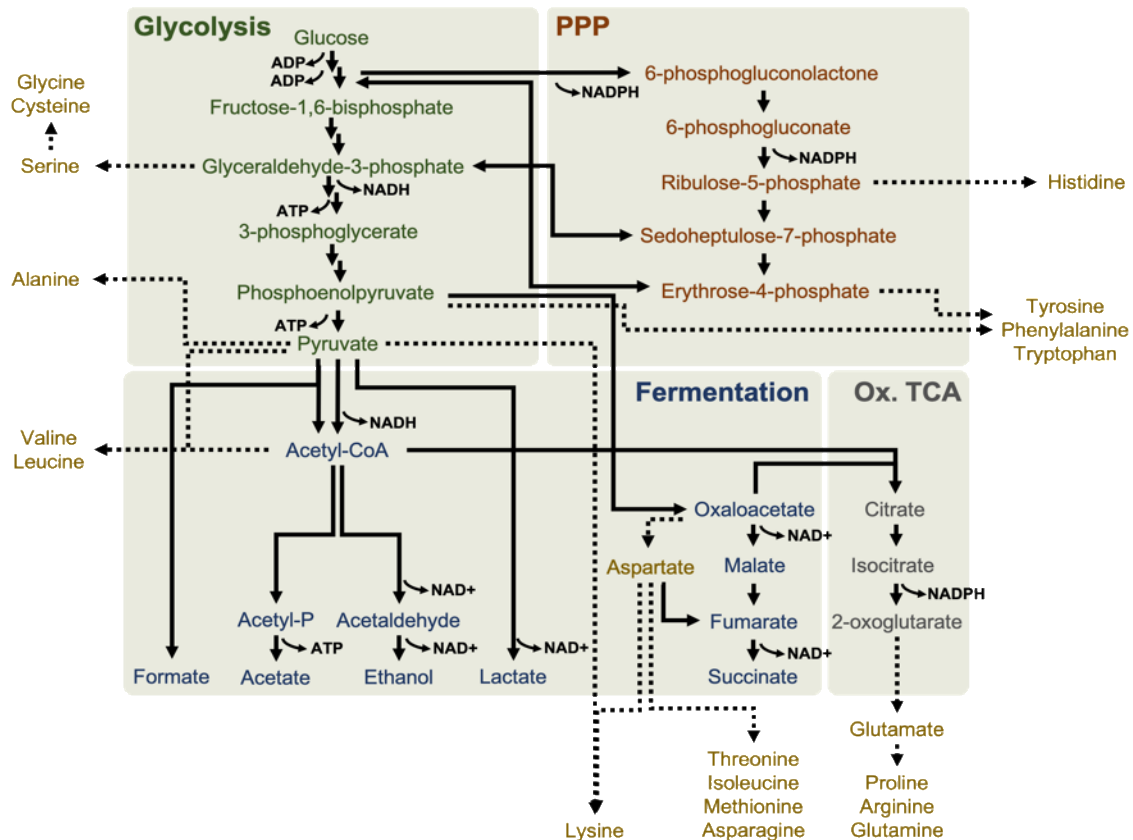


Figure 2.1. A putative metabolic model of lysates derived from *E. coli* BL21DE3-Star growing exponentially in high glucose concentrations. Intermediates and end-products of glycolysis (green), the pentose phosphate pathway (dark orange), and fermentative pathways (blue) from acetyl-CoA have been reported in lysate-based CFME. The presence of succinate fermentation implies the activation of the oxidative TCA branch (gray). Amino acid anabolism (gold) in lysates is not well-defined and is investigated here.

2.3. Methods

2.3.1. Starting, stopping, and processing time course CFME reactions for HPLC-RID quantification

1. Thaw previously prepared *E. coli* lysates and prepare the rest of the reaction components on ice. NOTE: The lysates reported here were derived from *E. coli* BL21DE3-Star grown in 2xYPTG (1.8 % glucose) media to mid-log phase.
 - 1.1. Prepare an appropriate volume of filter-sterilized (0.20 μ m pore filter) S30 buffer (1 M Tris-OAc adjusted to pH 8.2 with glacial acetic acid, 1.4 M Mg(OAc)₂, and 6 M KOAc).
 - 1.2. Prepare an energy mix containing glucose, glutamate salts, ATP, Coenzyme A, NAD⁺, Bis-Tris buffer, and dipotassium phosphate in the S30 buffer. Final concentrations in the desired reaction volume used to prepare CFME reactions here were 100 mM glucose, 18 mM magnesium glutamate, 15 mM ammonium glutamate, 195 mM potassium glutamate, 1 mM ATP, 0.2 mM Coenzyme A, 1 mM NAD⁺, 150 mM Bis-Tris, and 10 mM dipotassium phosphate.
2. Combine the components in 1.5 mL microcentrifuge tubes to prepare final reactions with 4.5 mg/mL of total lysate protein. Here, CFME reactions were prepared with final volumes of 50 μ L in triplicate per timepoint. Incubate the reactions at 37 °C for their respective timeframes. NOTE: Work fast and add lysate as the final component of the reaction mix to prevent premature metabolic reactions with glucose and glutamate salts. Minimal glucose consumption can occur depending on how long reaction mixes are incubated on ice.
3. Terminate the reactions and process the samples for HPLC-RID analysis.
 - 3.1. To terminate the triplicate reactions at their appropriate time points, immediately add an equal volume of 5% trichloroacetic acid to each sample's final reaction volume (i.e., 50 μ L of 5% trichloroacetic acid to a 50 μ L reaction). Dilute each sample with sterile water at 2x the reaction volume (i.e., 100 μ L).

- 3.2. To recapitulate time zero, mix the same volume of 5% trichloroacetic acid as the total final reaction volume (i.e., 50 μ L) with the lysate prior to adding the rest of the reaction components. This acidification step precipitates lysate enzymes before they significantly metabolize glucose.
- 3.3. Vortex the samples and centrifuge on a benchtop microcentrifuge at 11,600 $\times g$ for 5 min and transfer supernatants containing the organic analytes to clean tubes. Store the samples at -20 $^{\circ}$ C if HPLC analyses are to be conducted on a different day. Ensure to thaw the stored samples on ice before proceeding to the next step.
- 3.4. Filter each supernatant with a 0.22 μ m pore filter. As an alternative to syringes, use centrifuge tube filters and centrifuge the supernatants at 16,300 $\times g$ for 1 min.
- 3.5. Transfer each filtrate to a clean HPLC glass vial. Load vials onto the HPLC autosampler tray.
4. Prepare samples for standard curve generation.
 - 4.1. Prepare a stock solution of all target analytes dissolved in S30 buffer at equimolar amounts above the starting concentration of glucose in the CFME reactions. Here, a stock solution consisting of 150 μ M glucose, succinate, lactate, formate, acetate, and ethanol, was prepared. Perform 1:1 (v/v) serial dilutions from the stock solution to obtain triplicate 50 μ L solutions with final concentrations ranging from 0 μ M to the stock concentration (i.e., 150 μ M).
 - 4.2. Dilute each solution with 50 μ L of 5% trichloroacetic acid and 100 μ L of sterile water. Repeat steps 3.4–3.5. NOTE: It is good practice to run solutions for standard curve generation with each batch of samples to ensure accurate quantification of metabolite concentrations.

2.3.2. Preparing the HPLC system for metabolite detection

1. Under a fume hood, prepare a sterilized 5 mM sulfuric acid solution from deionized and filter-sterilized water. Add ~550 μL of a 98% HPLC grade sulfuric acid solution to 2 L of water to prepare 5 mM sulfuric acid. CAUTION: Sulfuric acid is a hazardous chemical and working under a fume hood with proper lab PPE prevents inhalation, skin contact, and eye contact. Concentrated sulfuric acid reacts vigorously with water and should be added directly to water, not the other way around. Store in a cool, dry area away from direct sunlight and follow proper waste disposal measures set by the laboratory.
2. Keep the 2 L bottle of 5 mM sulfuric acid incubated in a water bath next to the HPLC instrument. Set the water bath to 35 °C. Place tubing with a solvent filter in the solvent bottle and attach the other end to a degasser module in line with the pump module. NOTE: Purging the system with a freshly prepared solvent prior to installing the column is good instrument handling practice.
3. Equip the HPLC instrument with the HPLC column in line with the RID module. Place the column in the 35 °C water bath if a column thermostat is not available.
4. Prepare the RID module for analysis at 35 °C in OpenLab CDS (Online) 2.15.26 software installed on the system computer.
 - 4.1. In the View menu, select the Method and run Control view. Right-click on the Pump Module > Method. Set the flow rate to 0.55 $\text{mL}\cdot\text{min}^{-1}$ and select the On button to initiate the pump. NOTE: If the column was in storage prior to being equipped on the HPLC, ramp up the flow rate to 0.55 $\text{mL}\cdot\text{min}^{-1}$ after equilibrating the column following the manufacturer's instructions.
 - 4.2. Right-click on the panel corresponding to the RID Module > Method. Set the temperature of the RI detector module to 35 °C and select On to start warming up the RI detector module.

4.3. Right-click on the panel RID Module > Control. Select On for the Purge Reference Cell for at least 15 min when using a fresh solvent or 1 h if different solvents flowed through the RI detector prior to this setup. Click on the On button. NOTE: Keep the pump and RI detector on until a stable baseline is achieved on the online plot. This is affected by temperature fluctuations in the laboratory and can take up to 4 h or longer. It is recommended that the system is kept on overnight prior to sample loading.

2.3.3. Creating a method for the isocratic HPLC separation of organic fermentation products in OpenLab (Online)

1. From the menu bar, select Method > New Method. Select Method > Save Method as [MethodName].M. Select Method > Edit Entire Method > Instrument/Acquisition.
2. Within the Binary Pump tab, set the flow to $0.55 \text{ mL} \cdot \text{min}^{-1}$. Under Solvents, select the letter corresponding to the solvent input on the pump module and set it to 100% for isocratic elution. Set Pressure Limits to 0 and 400 bar and input 30 min as the Stoptime.
3. Within the Sampler tab, set the Injection Volume to 50 μL . Select the As Pump/No Limit option under Stoptime. Set the Advanced Auxillary Settings for Draw Speed, Eject Speed, and Draw Position to $200 \text{ } \mu\text{L} \cdot \text{min}^{-1}$, $200 \text{ } \mu\text{L} \cdot \text{min}^{-1}$, and -0.5 mm.
4. Within the RID tab, set Optical Unit Temperature to 35 °C. Under Signal, select Acquire for Signal and > 0.2 min for Peakwidth. Select the As Pump/Injector option for Stoptime.
5. Under Advanced within the RID tab, set Analog Output to 5% Zero Offset and 500,000 nRIU for Attenuation. Select the Positive option for Signal Polarity and the On option for Automatic Zero Before Analysis.
6. Save the method by selecting Method > Save Method. Ensure that the method is loaded by selecting Method > Load Method > [MethodName].M.

2.3.4. Creating a Sequence Table for autosampling and start the HPLC-RID system for data acquisition

1. From the menu bar, select Sequence > New Sequence Template. Select Sequence > Save Sequence Template as [SequenceTemplateName].S.
2. Select Sequence > Sequence Table. Append 'n' rows corresponding to 'n' vials, then input vial positions and sample names under Vial and Sample Name, respectively, according to their arrangement on the autosampler tray. Select the Method generated in step 3 from the Method Name dropdown menu and input 50 µL as Inj/Vial (Injection per Vial) for each row.
3. Click Apply and save the sequence template by selecting Sequence Template > Save Sequence Template. Ensure the sequence template is loaded by selecting Sequence > Load Sequence Template > [SequenceTemplateName].S.
4. Once a stable baseline is achieved on the online plot, right-click the panel RID Module > Control > Off Recycling Valve to ensure that solvent flow through the RID detector is directed to waste. To start data acquisition, select Sequence from the menu bar, Sequence > Run.

2.3.5. Extracting and analyzing data post-run

1. Select Data Analysis view from the View menu. Locate the Sequence File name from the File List on the left-hand side of the screen. On the center panel on the screen, go to the Signal View Selection > RID Signal to view the sample chromatograms.
2. Select a row corresponding to a high concentration standard sample from the top panel on the screen. Take note of retention times for the target analyte peaks on the displayed chromatogram. Peaks corresponding to the target analytes will be arranged along the retention time axis as glucose, succinate, lactate, formate, acetate, and ethanol (**Appendix Figure 2.7**). NOTE: The first large peak on the chromatogram corresponds to trichloroacetic acid. Its RI units should be consistent across all standard curve samples. Validate the retention time of each target analyte by running each compound as a separate sample.

3. Extract peak areas for each target analyte from chromatograms of standards and reaction samples.
 - 3.1. Discern whether peaks-of-interest are well integrated by the software. Draw the red line as the base of each peak to obtain an accurately integrated area under the curve. If automatic integration fails (i.e., red line is askew), select the Manual Integration button from the Integration Tool Set and manually draw a peak base to integrate the peak area.
NOTE: If manual integration must be performed for a target analyte in one sample, keep consistent and manually integrate the same analyte across all samples.
 - 3.2. Select the Cursor tool from the Common Tool Set to click on properly integrated peaks. The peak area and corresponding reaction time of the select peak will be highlighted as a table row on the bottom panel of the screen.
 - 3.3. To export peak areas, select File > Export > Integration Results.
4. Quantify the target analyte concentrations using standard curves.
 - 4.1. Plot peak area values vs. known concentrations of samples (spreadsheet). Right-click on the plotted data, Add Trendline > Format Trendline > Display Equation on Chart.
 - 4.2. In a separate spreadsheet, use the equations of standard curve trendlines to convert peak area values to concentrations for every analyte from each sample. Calculate average peak areas and standard error values across triplicates for data visualization.

2.3.6. Starting, stopping, and processing time course isotope tracing CFME reactions for LC-MS/MS quantification

1. Set up triplicate reactions per time point (except time zero) on ice as described in 1.1–1.2. However, instead of glucose, use a final concentration of 100 mM $^{13}\text{C}_6$ -glucose in the reactions. Here, the reactions were incubated at 37 °C for 1 h, 2 h, and 3 h.
2. To terminate, flash freeze the reactions in liquid nitrogen and store at -80 °C. This storage step can be skipped for same-day analysis. NOTE: Trichloroacetic acid was not used to stop

reactions due to interference from the acid when detecting some central carbon metabolites via LC-MS/MS. Instead, extraction solvent containing formic acid (step 6.3) is used to precipitate metabolic proteins since formic acid's mass is below the detection limit of the reported MS/MS method.

3. Prepare 50 mL of extraction solvent. Combine and vortex 20 mL of acetonitrile, 20 mL of methanol, and 10 mL of water (all LC-MS grade) in a 50 mL centrifuge tube along with 0.199 mL of formic acid to make a 0.1 M solution. Chill the solvent to 4 °C during extraction and store the solvent at -20 °C when not in use.
4. Processing samples for LC-MS/MS analysis
 - 4.1. On the day of analysis, pipette an equivalent volume of extraction solvent (i.e., 50 µL) to each sample. If the samples were frozen, add the extraction solvent before samples completely thaw out to prevent the reactivation of glucose metabolism. Perform all sample processing steps on ice.
 - 4.2. To recapitulate time zero, pipette the final volume of extraction solvent (i.e., 50 µL) to an appropriate volume of lysate for the desired final concentration in the reaction (i.e., of 4.5 mg/mL in 50 µL reaction volume). Add the rest of the reaction components as in step 1.2. This acidification step precipitates lysate enzymes before they significantly metabolize glucose.
 - 4.3. Incubate the samples in extraction solvent on ice for 30 min with gentle shaking, then centrifuge the samples at 21,000 x g for 15 min at 4 °C to separate the supernatant from the precipitated protein. Transfer 50 µL of the supernatant to autosampler vials and load the vials onto the tray within the 4 °C autosampler. Store the rest of the supernatant at -20 °C for future analyses.

2.3.7. Setting up the LC system for LC-MS/MS analysis

1. Prepare 1 L of Solvent A by completely dissolving 77.08 mg of ammonium acetate in 950 mL of water and 50 mL of isopropanol. Prepare 1 L of Solvent B with 650 mL of acetonitrile, 300 mL of water, and 50 mL of isopropanol along with 77.08 mg of ammonium acetate. Ensure that all solvents are LC-MS grade.
2. Connect the solvent bottles containing Solvents A and B to the pump module. Purge the system at a high flow rate to remove/limit any air contamination that may have occurred during the equipment of the solvents to the LC system.
3. Equip the system with a C18 reversed-phased column (30 cm column length, 75 μ m inner diameter, and 5 μ m particle diameter). Condition the column to the LC-MS system by flowing 100% solvent B and slowly flowing up Solvent A to 100%. NOTE: Column tips were prepared in-house using a micropipette puller and packed with pressure cells and helium.

2.3.8. Creating a method on LC-MS/MS data acquisition and interpretation software for the LC system linked to Fourier Transform and Ion Trap Mass Spectrometers

1. Open the Tune Plus software to edit a tune file for the MS method.
 - 1.1. From File on the menu bar, open a preinstalled negative mode tune file.
 - 1.2. Select ScanMode on the menu bar, then select Define Scan Window. Set the microscan time setting for MSn to 1 for both Ion Trap and FT.
 - 1.3. Go to settings for Nano-ESI Source and set Spray Voltage to 4 kV. Modulate this until an acceptable electrospray is generated; typically, acceptable electrospray can be achieved within the range of 2–5 kV.
 - 1.4. Save the tune file.

2. Generate a new LC method using the Setup Wizard of the instrument's data acquisition and interpretation software. Open Roadmap > Sequence Setup > Wizard. Since these methods do not require the usage of a column heater, skip the Temp Control step.
 - 2.1. Under Flow Gradient Pump Options, select Multistep. In the next window, insert 7 lines and set the flow rate for each row to 0.1 mL·min⁻¹. Input the following parameters for each row: from 0–3 min, deliver 100% solvent A; from 3–9 min, introduce a gradient from 100% solvent A to 20% solvent B; from 9–19 min, introduce a new gradient from 20% solvent B to 100% solvent B; from 19–27 min, hold at 100% solvent B; from 27–28 min, set the gradient back to 100% solvent A; from 28–44 min, rinse and recondition of the column for subsequent runs by holding at 100% solvent A. Include a final step to lower the flow rate to 0.03 mL·min⁻¹ upon completion of the run to conserve the solvent when the LC is not in use.
 - 2.2. Apply Default settings for Sampler Options > Pump Pressure as the Acquisition Option and use Default Acquisition Time and use Default pump pressure options.
3. Create an MS/MS method by selecting the Orbitrap Velos Pro MS icon from the sidebar on the Instrument Setup window.
 - 3.1. Click on New Method > Data Dependent MS/MS. Set Acquire Time to the length of the LC run (i.e., 44 min), Segment to 1, and Scan Events to 11. For Tune File, select the edited file from step 8.1. NOTE: The first event is an MS1 precursor scan using the Fourier Transform MS (FTMS). The succeeding 10 events will be MS2 scans selecting the 10 most intense and unique ions in each precursor scan for MS2 fragmentation.
 - 3.2. For Event 1, under Scan Description set Analyzer to FTMS and Polarity to Negative. Under MSn Settings, use a Resolution of 30,000 and a Normalized Collision Energy of 35 V. Set Scan Ranges to 50 m/z for first mass and 1800 m/z for last mass to capture small molecules.

- 3.3. For events 2 through 11, under Scan Description set Analyzer to Ion Trap. Select Dependent Scan and click on Settings > Global > Dynamic Exclusion and select Enable; set a 30 s repeat duration and 120 s exclusion duration to eliminate repeat scans in proximity.
- 3.4. Go to Scan Event settings and set Mass Determined from Scan Event to 1 for all MS2 events (2 through 11). To scan for the top 10 most intense ions, set each MS2 scan event to detect an nth Most Intense Ion from 1st to 10th. Therefore, set Event 2 to detect 1 as the nth Most Intense Ion, event 3 to detect 2, and so on.
- 3.5. Close the setup window and go to File | Save As [Method_Name].meth. NOTE: For general use, maintenance, and calibration of the LC instrument and mass spectrometer, refer to the operating instructions and manuals supplied by the manufacturer.

2.3.9. Setting up a run sequence and starting the LC-MS/MS run

1. Set up a Run Sequence using the LC-MS/MS system's data acquisition and interpretation software. Within Roadmap > Sequence Setup, right-click on the table to insert as many rows as samples. For each row, set the Inj Vol to 5 µL and Position to the vial's respective position on the autosampler tray. Input file names as sample names and set the desired file path for run results. NOTE: Blank vials containing Solvent A can be run at the start of the sequence and between each set of triplicate samples (each set of time points) to rinse the column.
2. To start the run, highlight all File Names in the Sequence. From the menu bar, select Actions > Run Sequence > OK.

2.3.10. Consolidating files and searching for tentative annotations on MZmine 2.53

1. Open MZmine and import the '.raw' output files from step 9.1. From the menu bar, Raw Data Methods > Raw Data Import. Select files corresponding to samples.

2. Construct a list of peaks distinguishing between MS1 and MS2 scans. From the menu bar, Raw Data Methods > Feature Detection > MS/MS Peaklist Builder. Relevant settings include m/z Window set to 0.01 and Time Window set to the length of the run. Under set filters, select Negative as Polarity and Centroided as the Spectrum Type.
3. From the menu bar, go to Feature List Methods > Feature Detection > Peak Extender. Set m/z Tolerance to 0.005 m/z or 10 ppm and Minimum Height to 1E3. This step will create fully fleshed out peaks.
4. Remove duplicate peaks. Go back to Feature List Methods > Filtering > Duplicate Peak Filter. Relevant settings include m/z Tolerance set to 0.005 m/z or 10 ppm and a RT Tolerance set to 5 min.
5. To align peaks within similar data files (i.e., those of triplicate reactions), go back to Feature List Methods > Normalization > Retention Time Calibration. Be sure to process triplicate samples together and leave out blanks. Relevant settings include m/z Tolerance set to 0.005 m/z or 10 ppm, RT Tolerance set to 3 min absolute (min), and Minimum Standard Intensity set to 1E3.
6. Align peaks from all files by m/z and retention time from Feature List Methods > Alignment > RANSAC Aligner. Set m/z Tolerance to 0.005 m/z or 10 ppm, RT Tolerance and RT Tolerance After Correction to 44 and 39 min, respectively, RANSAC Iterations to 0, Minimum Number of Points to 10%, and Threshold Value to 1. Tick the option Require Same Charge State.
7. Correct for any data points that may have been lost in prior steps in Feature List Methods > Gap Filling > Peak Finder. Relevant settings include Intensity Tolerance set to 50%, m/z Tolerance set to 0.005 m/z or 10 ppm, and RT tolerance set to 3 min. Enable RT correction.

2.3.11. Calculating negative mode masses of ¹³C-labeled glucose-derived metabolites and searching for the m/z features of these analytes in filtered data

1. Calculate the masses of ¹³C-labeled metabolites from glucose metabolism for the targeted search.
 - 1.1. Calculate the monoisotopic mass of each target compound from the number of atoms in the compound's molecular formula and the monoisotopic masses of each element.
 - 1.2. Calculate the compound's negative mode mass [M-H]⁻ by subtracting the mass of 1 proton (1.007276 Da) from the monoisotopic mass. This is the mass detected by negative mode MS detection after molecules are stripped of a hydrogen ion during ionization.
 - 1.3. From the negative mode mass, calculate the mass of the ¹³C-incorporating metabolite. Here, the masses of isotopologues that maximally incorporate glucose-derived ¹³C-labels were calculated.
2. Use calculated masses of ¹³C-labeled metabolites to search and annotate m/z features from MZmine results. For each possible hit, calculate the mass error (ppm) using the following equation:

$$ppm\ error = \frac{(\frac{m}{z})_{theoretical} - (\frac{m}{z})_{experimental}}{(\frac{m}{z})_{theoretical}} \times 10^6$$

NOTE: Experimental m/z values with < 15 ppm mass error were considered as putative annotations in the current analysis.

3. Manually check spectra of putative annotations on a quality browser to confirm annotations.
 - 3.1. Open Roadmap > Qual Browser. From the Tool Bar, Open Raw File to import raw MS data of each sample.
 - 3.2. Draw a line under a desired range of retention times (i.e., corresponding to the putative annotation) on the total ion chromatogram (top panel) to view a mass spectrum (bottom panel). Right click on the spectrum and input a range of masses that encompass the

target analyte's m/z. Check whether the putative annotations have distinct peak signals that are appreciably above noise (**Appendix Figure 2.8**).

4. Calculate the average peak areas and standard errors of positive annotations across biological replicates for each time point. Visualize the data (i.e., on a bar graph) to observe trends in glucose metabolism.

2.4. Representative Results

2.4.1. HPLC-RID based analysis of lysate fermentative metabolism

To quantify the lysate-based cell-free synthesis of common fermentation products from glucose, lysates derived from strains grown in 2xYPTG media were fed 100 μM glucose as a primary carbon source⁷⁴. Reactions were stopped over a 24 h time course by protein acidification. Filtered supernatants containing pyruvate, succinate, lactate, formate, acetate, and ethanol produced from glucose catabolism were loaded onto the autosampler module of a 1260 Infinity HPLC system equipped with a RID module. Vials with filtered mixtures of fermentative end-products and glucose at 1.17 μM , 2.34 μM , 4.69 μM , 9.38 μM , 18.75 μM , 37.50 μM , 75 μM , and 150 μM concentrations in S30 buffer were loaded onto the instrument as standards. Analytes were eluted isocratically from an HPLC column to the RID. Peaks for glucose, succinate, lactate, formate, acetate, and ethanol within the 1 to 150 μM range could be resolved by RID. Peak areas for glucose were derived by manual integration from the RID data for time course and standard curve samples. Extracted peak areas for succinate, lactate, formate, acetate, and ethanol were taken from automatically integrated signals. All standard curves (peak area vs. known concentration) had R^2 values >0.99 and were linear throughout the range of concentrations used here.

Molar concentrations for all target analytes were calculated from their respective standard curves. Glucose was consumed within the first 3 h of the reaction and mainly fermented to lactate

(Figures 2.2A & B). Ethanol accumulation also significantly occurred within the first 3 h of the reaction and stopped thereafter **(Figure 2.2C)**. The observation of significant lactate and ethanol production with significant glucose consumption after 3 h was not unprecedented since lactate and ethanol production pathways allow the regeneration of 1 net mol NAD⁺ from glycolytic NADH required for continued glucose consumption through glycolysis **(Figure 2.1)**. Lactate and ethanol can thus be considered as the major fermentation end-products in lysate-based cell-free glucose metabolism. Acetate was initially present in the reactions as a component of the S30 buffer and unexpectedly only accumulated as a result of metabolism after 6 h when glucose consumption had slowed down **(Figure 2.2D)**. This result suggests that acetate fermentation does not necessarily enable rapid glycolytic flux in earlier time points. Meanwhile, formate and succinate were synthesized as minor fermentation products **(Figures 2.2E & F)**. Altogether, the method enabled the absolute quantification of sugar substrate depletion and fermentative product formation in *E. coli* S30 lysates.

2.4.2. Tracing carbon metabolism in lysates via LC-MS/MS

MS detection to profile lysate glucose metabolism specifically was applied here. Lysates derived from strains grown in 2xYPTG media were fed ¹³C₆-glucose as a carbon source. CFME reactions were run in triplicate for 0 h, 1 h, 2 h, and 3 h. Samples from each timepoint were loaded on an LC system equipped with a reversed-phase column and coupled to Fourier transform and ion trap mass spectrometers. Negative ion mode spectra were obtained and processed to analyze organic acids, sugar phosphates, and amino acids. Calculated theoretical masses of ¹³C-labeled species belonging to central carbon metabolism were searched to identify compounds that are specifically glucose-derived. Based on the utilized source strain cultivation conditions and previous reports of active pathways in *E. coli* CFME, it is assumed here that the lysate proteome comprises a metabolic network that feeds glucose into glycolytic fermentation, the pentose phosphate pathway, and possibly amino acid anabolism^{46,72–74,78} **(Figure 2.1)**. The search was

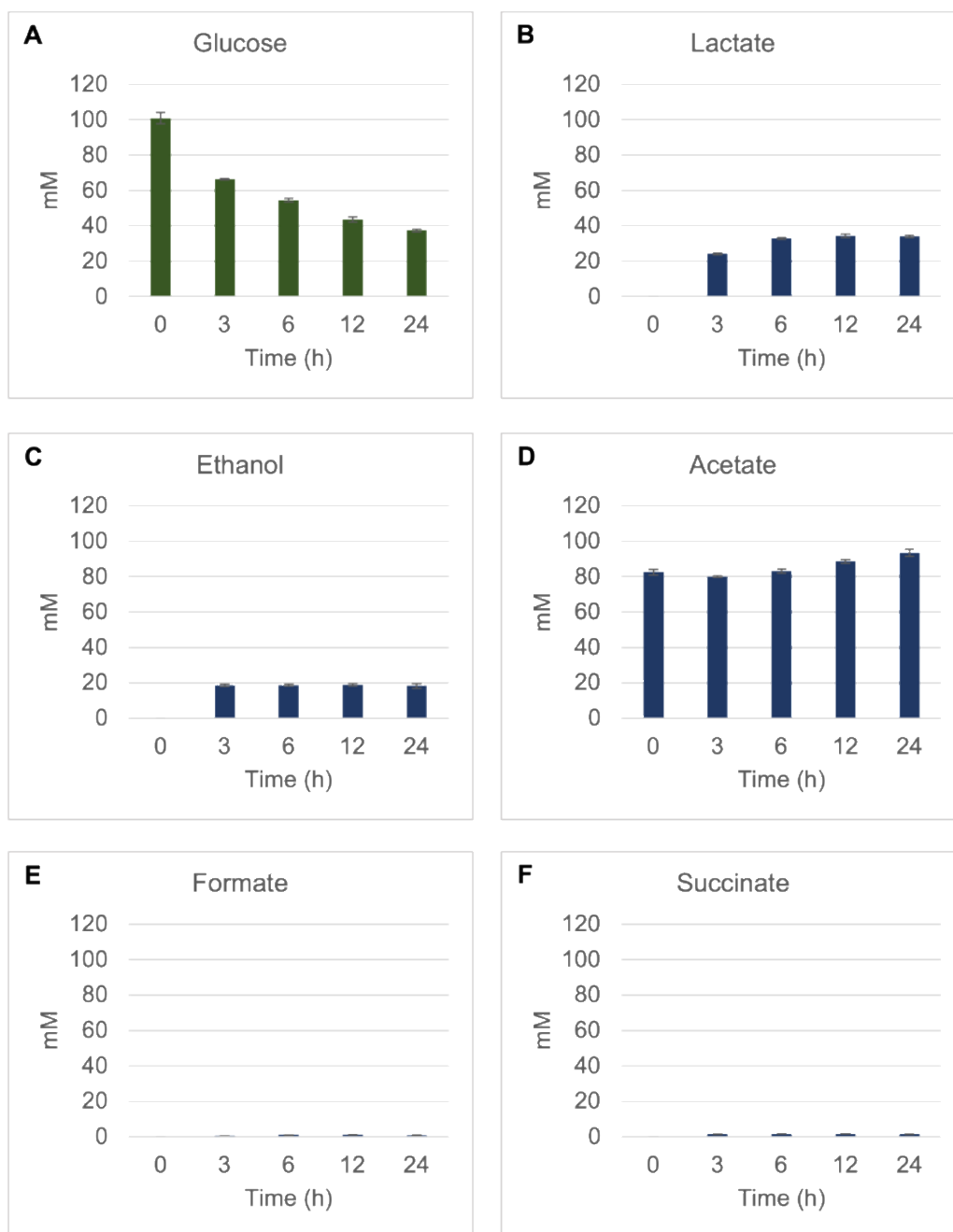


Figure 2.2. HPLC-RID data for glucose consumption and fermentative end-product synthesis in CFME reactions prepared with *E. coli* crude extracts. (A) Glucose consumption and (B) lactate, (C) ethanol, (D) acetate, (E) formate, and (F) succinate production in CFME reactions were monitored over 24 h. Average mM concentrations and error bars (SE) quantified with standard curves are presented. (n = 3)

therefore narrowed down to members of these pathways of which 16 metabolites incorporating glucose-derived ^{13}C labels were unambiguously annotated (**Appendix Table 2.1**).

$^{13}\text{C}_6$ -glucose was observably consumed through glycolysis as evidenced by the fluctuations in glycolytic intermediate abundances (**Figures 2.3A to E**). Consistent with the HPLC-RID data, glucose accumulated to $^{13}\text{C}_3$ -lactate and was also fermented to $^{13}\text{C}_3$ -succinate within the first 3 h of the reaction (**Figures 2.4A & B**). The formation of $^{13}\text{C}_3$ -succinate isotopologue supports the proposed model of lysate glucose metabolism (**Figure 2.1**) where succinate is likely to be generated by the carboxylation of a 3-carbon phosphoenolpyruvate (PEP) molecule, and not from the entry of a 2-carbon acetyl-CoA molecule to the TCA cycle. Activation of the TCA cycle has been assumed in previous CFME studies, but other ^{13}C -labeled intermediates of TCA were not detected here^{49,53,74}. $^{13}\text{C}_3$ -aspartate synthesis however occurred within the first h and was consumed, reinforcing the idea that PEP is directly converted to oxaloacetate (**Figures 2.1 & 2.5C**).

The data are reflective of a lysate proteome from source strains harvested during fermentative growth on glucose-rich media (2xYPTG). This would further imply that the rest of the TCA enzymes not participating in succinate production form an oxidative TCA branch (**Figure 2.1**). None of the metabolites in this pathway however were detected, possibly because high concentrations of glutamate added to the CFME reaction as a salt counterion prevent the progression of this branch.

The HPLC-RID data is additionally complemented by the lack of $^{13}\text{C}_2$ -acetate detection within the 3 h reaction timeframe suggesting no build-up of acetate from glucose up to 3 h (**Figure 2.2B**). However, the direct precursor of acetate, acetyl-phosphate (acetyl-P), accumulated suggesting that the Pta arm of the Pta-AckA pathway for acetate synthesis from acetyl-CoA is active (**Figures 2.4C & D**). The AckA catalyzed dephosphorylation of $^{13}\text{C}_2$ -acetyl-P to $^{13}\text{C}_2$ -acetate likely does not occur within this timeframe due to acetate being a major component of S30 buffer used in the reactions (**Figures 2.1 & 2.2B**).

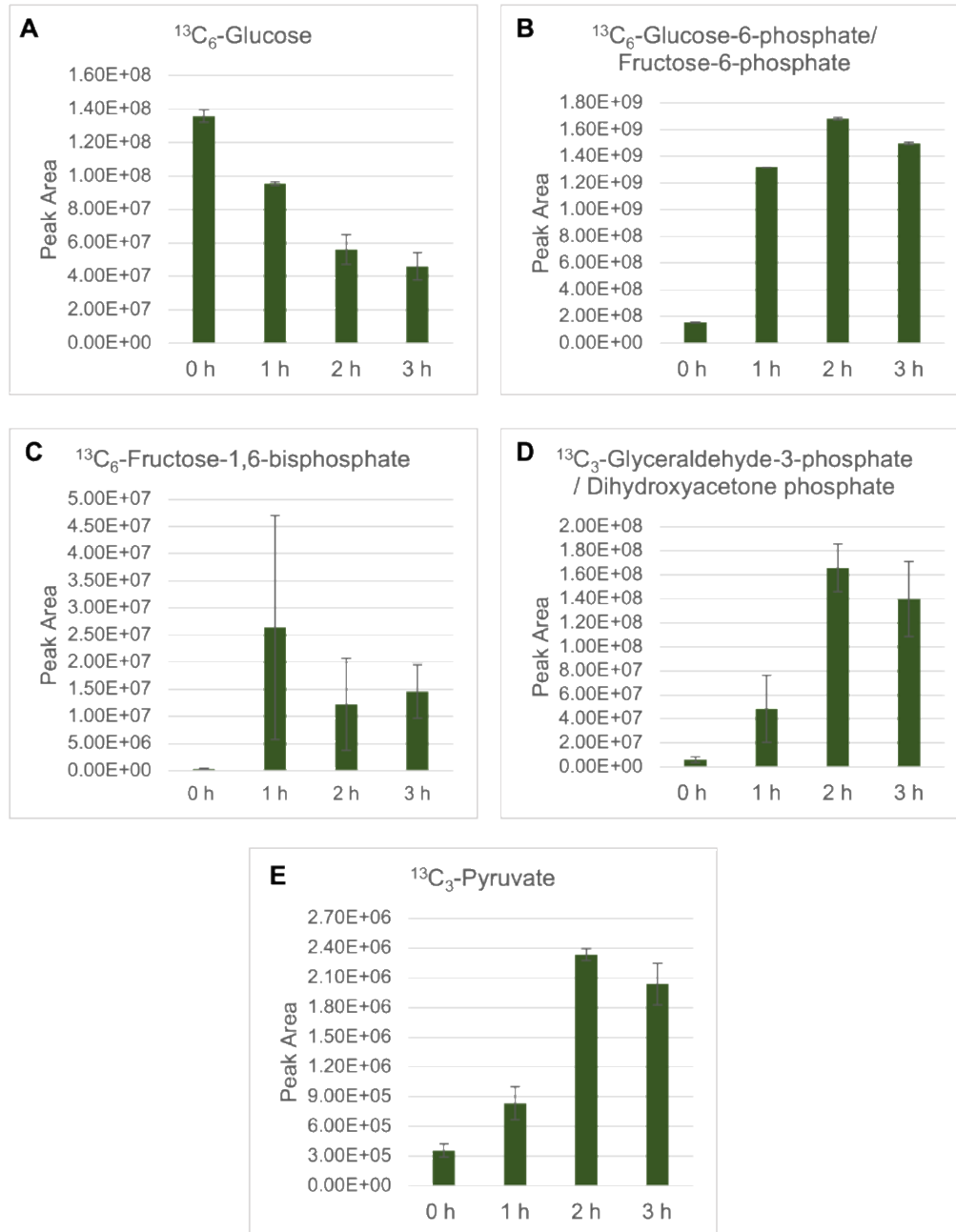


Figure 2.3. Time course trends of $^{13}\text{C}_6$ -glucose and ^{13}C -labeled glycolytic intermediates in *E. coli* lysate CFME. Relative abundances of (A) $^{13}\text{C}_6$ -glucose, (B) $^{13}\text{C}_6$ -glucose-6-phosphate/fructose-6-phosphate, (C) $^{13}\text{C}_6$ -fructose-1,6-bisphosphate, (D) $^{13}\text{C}_3$ -glyceraldehyde-3-phosphate/dihydroxyacetone phosphate, and (E) $^{13}\text{C}_6$ -pyruvate in CFME reactions over 3 h. Raw peak areas extracted by mzMINE software were used to calculate averages and error bars (SE) for positive annotations. (n = 3)

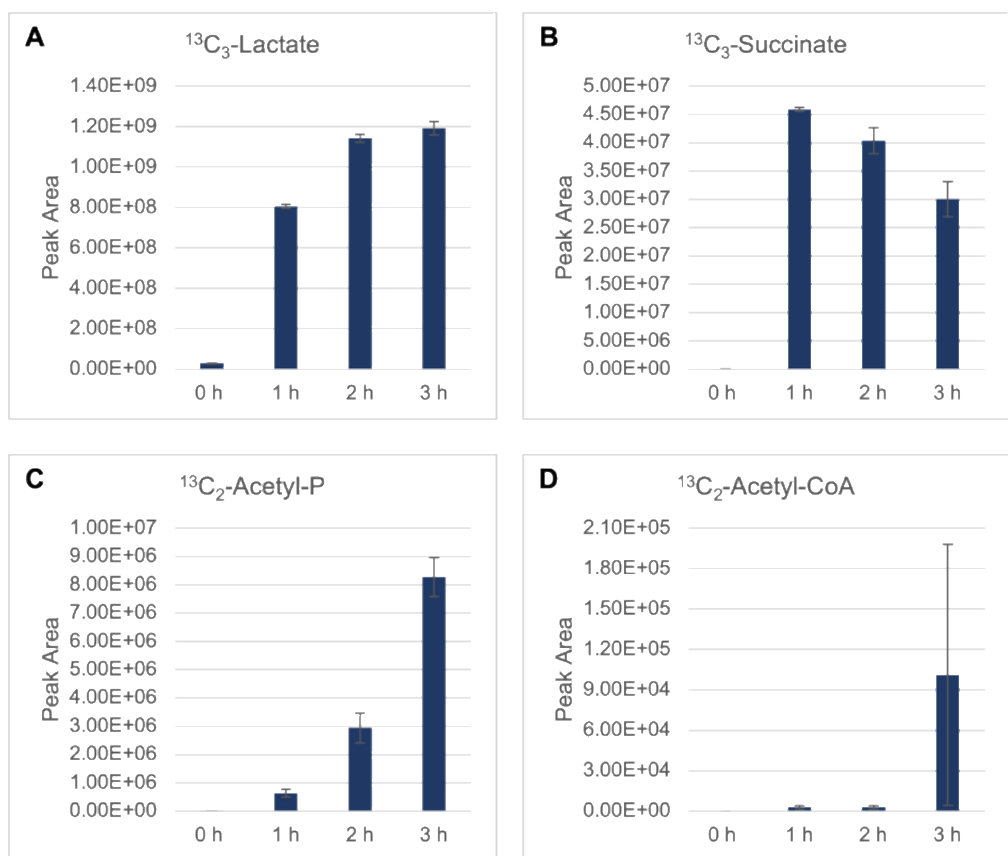


Figure 2.4. Time course trends of intermediates and end-products in $^{13}\text{C}_6$ -glucose fermentation in *E. coli* lysate CFME. Relative abundances of (A) $^{13}\text{C}_3$ -lactate, (B) $^{13}\text{C}_3$ -succinate, (C) $^{13}\text{C}_2$ -acetyl-phosphate, and (D) $^{13}\text{C}_2$ -acetyl-CoA in CFME reactions over 3 h. Raw peak areas extracted by mzMINE software were used to calculate averages and error bars (SE) for positive annotations. (n = 3)

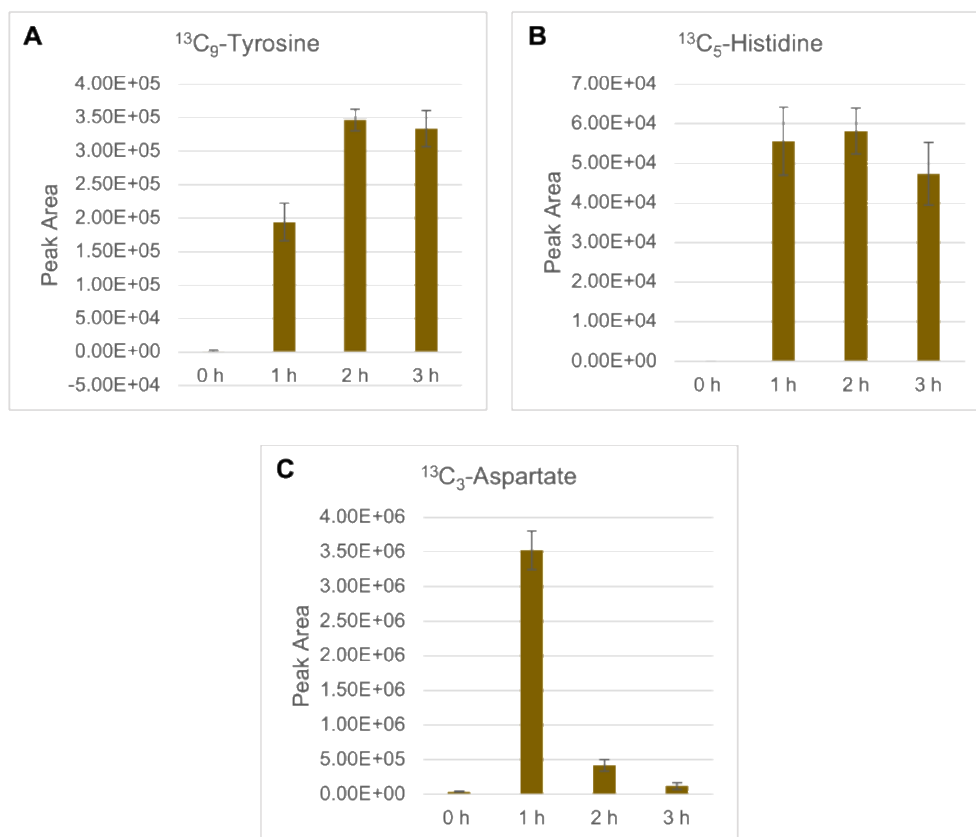


Figure 2.5. Time course trends of detected $^{13}\text{C}_6$ -glucose derived amino acids in *E. coli* lysate CFME. Relative abundances of (A) $^{13}\text{C}_9$ -tyrosine, (B) $^{13}\text{C}_5$ -histidine, and (C) $^{13}\text{C}_3$ -aspartate over 3 h. Raw peak areas extracted by mzMINE software were used to calculate averages and error bars (SE) for positive annotations. (n = 3)

The incorporation of $^{13}\text{C}_6$ -glucose derived carbons to sugar phosphates 6-phosphogluconolactone (6PGL), 6-phosphogluconate (6PG), ribulose-5-phosphate (Ru5P), and sedoheptulose-7-phosphate (S7P) was also observed (**Figure 2.6**). These results confirm the participation of the pentose phosphate pathway in lysate glucose metabolism and likely feeds $^{13}\text{C}_9$ -tyrosine synthesis, which has been suggested before by a proteomic study, while also providing a precursor for $^{13}\text{C}_5$ -histidine production⁷³ (**Figures 2.5A & B**). Labeled phenylalanine and tryptophan were not observed here and neither were most of the essential amino acids. However, this is not entirely surprising since amino acid anabolism is likely to be enriched in lysates derived from cells grown in nutrient-starved conditions or at the stationary phase^{73,82}. Moreover, the data thus far suggests that intermediates of glycolysis and fermentation are funneled towards cofactor regenerating end reactions, which must preclude the synthesis of many amino acids derived from glyceraldehyde-3-phosphate, pyruvate, and acetyl-CoA (i.e., glycine, cysteine, serine, alanine, valine, leucine, and lysine) (**Figure 2.1**).

As mentioned, $^{13}\text{C}_3$ -aspartate was produced within the first hour, whereas aspartate derived ^{13}C -incorporating amino acids (i.e., threonine, isoleucine, methionine, and asparagine) were not observed possibly because glucose-derived aspartate participates in fermentation (**Figures 2.1 & 2.5C**). Lastly, flux towards labeled glutamate and amino acids derived from glutamate may have been impeded by high levels of glutamate in the reaction environment (**Figure 2.1**).

2.5. Discussion

The outlined HPLC-RID approach can be used to successfully quantify sugar substrate consumption and subsequent conversions to major organic acid and alcohol products of lysate central metabolism over time. This protocol furthermore employs a simple isocratic method using a single mobile phase, requires minimal sample preparation, and allows a simple targeted

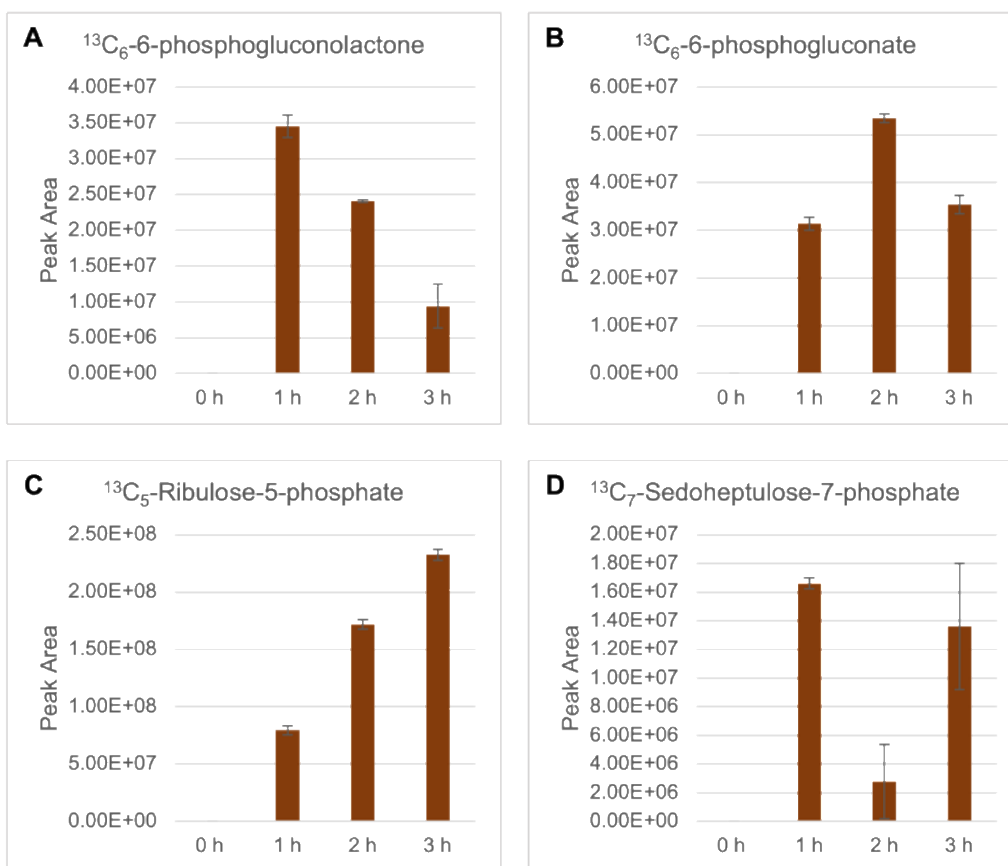


Figure 2.6. Time course trends of $^{13}\text{C}_6$ -glucose derived pentose phosphate pathway intermediates in *E. coli* lysate CFME. Relative abundances of (A) $^{13}\text{C}_6$ -6-phosphogluconolactone, (B) $^{13}\text{C}_6$ -6-phosphogluconate, (C) $^{13}\text{C}_5$ -ribulose-5-phosphate, and (D) $^{13}\text{C}_7$ -sedoheptulose-7-phosphate over 3 h. Raw peak areas extracted by mzMINE software were used to calculate averages and error bars (SE) for positive annotations. (n = 3)

downstream analysis. Analytes measured by the HPLC-RID method are distinguished solely by their retention times, and therefore their interactions with the selected column resin. The HPLC column utilized here was particularly designed for separating carbohydrates, organic acids, and alcohols by a combination of size-exclusion and ligand exchange (i.e., ion-moderated partition chromatography). The described method is, therefore, useful for more targeted analysis of carbohydrate substrates and select end-products of glucose fermentation pathways which are expected to primarily facilitate and energize lysate-based bio-transformations^{47,53,74}. However, this protocol does not account for the activation of other metabolic pathways in cell extracts. Pipelines employing other chromatographic separation techniques (i.e., hydrophilic interaction chromatography), gradient elution methods, more complicated sample preparation (i.e., derivatization), and different optical detectors (e.g., ultraviolet light or evaporative light scattering detectors) could be used to detect other metabolites such as amino acids and sugar phosphates^{83,84}. Alternatively, a global approach to studying lysate metabolism can be taken using LC-MS/MS.

The described LC-MS/MS method is a single workflow for measuring and identifying a broader range of metabolites. LC-MS/MS is a state-of-the-art analytical tool for metabolome profiling because of its sensitivity and ability to distinguish metabolites by retention time and m/z ratios with high resolution⁷⁹. With a focus on central carbon metabolic pathways and amino acid anabolism, negative mode MS/MS was implemented to specifically detect polar organic acids, sugar phosphates, and amino acids. Coupled with a nano-liquid chromatography technique, the method provides high sensitivity for detecting small molecules in the complex lysate background⁸⁰. In terms of profiling lysate-based CFME metabolism, however, a limitation of the described LC-MS/MS protocol is its lower detection limit of 50 m/z which precludes the measurement of ethanol, a major product in lysate glucose metabolism, as well as formate, which are both otherwise easily quantified by the detailed HPLC-RID method. Compared to LC-MS/MS, HPLC-RID has the additional advantage of relative accessibility in terms of cost and difficulty. To

the latter point, troubleshooting of the LC-MS/MS method described here may require some degree of expertise in mass spectrometry. Nonetheless, MS detection has uniquely appealing applications over RID as it can additionally distinguish labeled isotopes in metabolomes, an excellent technique for understanding carbon movement from supplemented substrates through the complex lysate metabolic network⁸¹. Such an approach was applied here by supplementing reactions with $^{13}\text{C}_6$ -glucose and analyzing the relative abundance values of downstream ^{13}C -incorporating metabolites. The analysis allowed the definition of active and inactive pathways, supporting previously reported assumptions and providing new insights about metabolic flux in lysates. Modifications can also be made within the method for specific analyses. For example, standard solutions of ^{13}C -labeled target compounds can be analyzed along with samples to achieve absolute quantitative measurements of glucose-derived molecules over time and make conclusions about flux distributions. Better detection of positively charged compounds can also be enabled within the current workflow by running sequences with .meth files adjusted for positive mode detection.

Analytical sampling in both described methods is conveniently automated, ensuring high reproducibility. Moreover, smooth analytical runs can be expected as long as proper instrument handling practices and maintenance are observed. When using these tools to analyze CFME reactions, more critical considerations should be made upstream and downstream of sampling. During sample preparation, it is important that time-course controls are representative of time-zero. Here, proteins were precipitated in lysates by acidification to stop metabolic reactions. For time-zero samples, the acid solvent was combined with lysate prior to adding the reaction mix containing glucose. Acidification with trichloroacetic acid effectively ensured that glucose is not metabolized at time-zero, as shown in the HPLC-RID data (**Figure 2.2**). While a similar procedure to quench glucose metabolism was performed in the reported LC-MS/MS analysis, ^{13}C -labeled metabolites were detected in time-zero samples, albeit at significantly low abundance values relative to samples extracted at later timepoints. Moreover, these observations were limited to

intermediates of glycolysis. The data suggest that the reactions retain some degree of glycolytic activity after acidification with the extraction solvent that is detected by this highly sensitive method. The extent of this activity, however, should be quantified. A previous study reported that acidic extraction solvents may not sufficiently quench intermediate glycolytic reactions but can stop significant glucose consumption⁴⁸. While this remains to be further investigated in the system used here, drastic changes in relative abundance values between time-zero and later timepoint samples can be interpreted as trends in glucose metabolism. Exploring alternative quenching methods, however is recommended in similar applications, specifically for obtaining absolute quantities of metabolic intermediates⁴⁸. Furthermore, good practices during downstream software analyses should also be observed. Consistency is imperative when manually integrating peak areas from RID signals to reduce human error. Manual integration should also be applied to peak areas of standards whenever manually integrated peak areas are being used to quantify metabolite concentrations in samples. Throughout the targeted LC-MS/MS analysis, tentative annotations from MZmine analysis should be validated by manual peak checking using an MS quality browser and m/z features should only be annotated when calculated mass errors are acceptable. Here, these analyses were performed manually for a limited set of targets since comprehensive and robust software for isotope searching are not yet established. However, such automated methods for searching ¹³C labeled metabolites are currently emerging and would streamline more complicated analyses as well like profiling lysates beyond central carbon metabolism⁸⁵. Advanced liquid chromatography is a robust and widely applied method for separating small molecules in complex metabolic mixtures⁷⁵. The described methods couple this separation technique with refractive index or mass spectrometric detection to successfully analyze metabolite conversions in lysate-based CFME reactions. HPLC-RID and LC-MS/MS are individually powerful tools for profiling active lysate metabolism and their complementarity can further be leveraged to address each technique's inherent limitations. The reported methods enable the application and development of CFME as they can be utilized towards understanding

lysate metabolism, monitoring improvements in targeted metabolite conversions, and elucidating alterations to metabolite flux when manipulating lysate metabolism.

Appendix

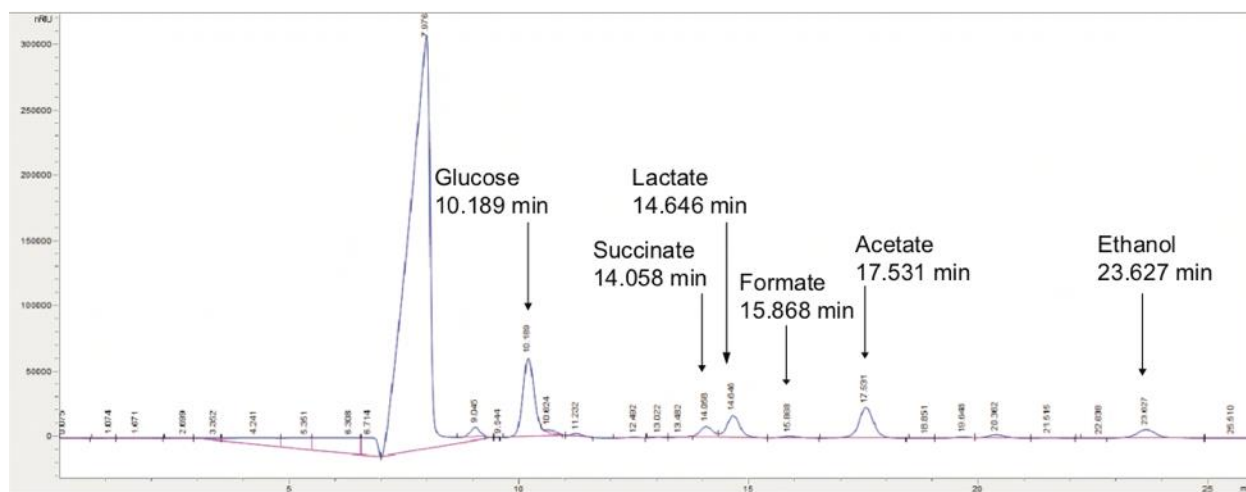


Figure 2.7. Representative HPLC-RID chromatogram showing peaks for major fermentative products in a CFME reaction incubated at 37 °C for 24 hrs. Glucose, succinate, lactate, formate, acetate, and ethanol peaks are sufficiently distinguishable by their retention times on an HPLC column during isocratic elution with 5 mM sulfuric acid solvent.

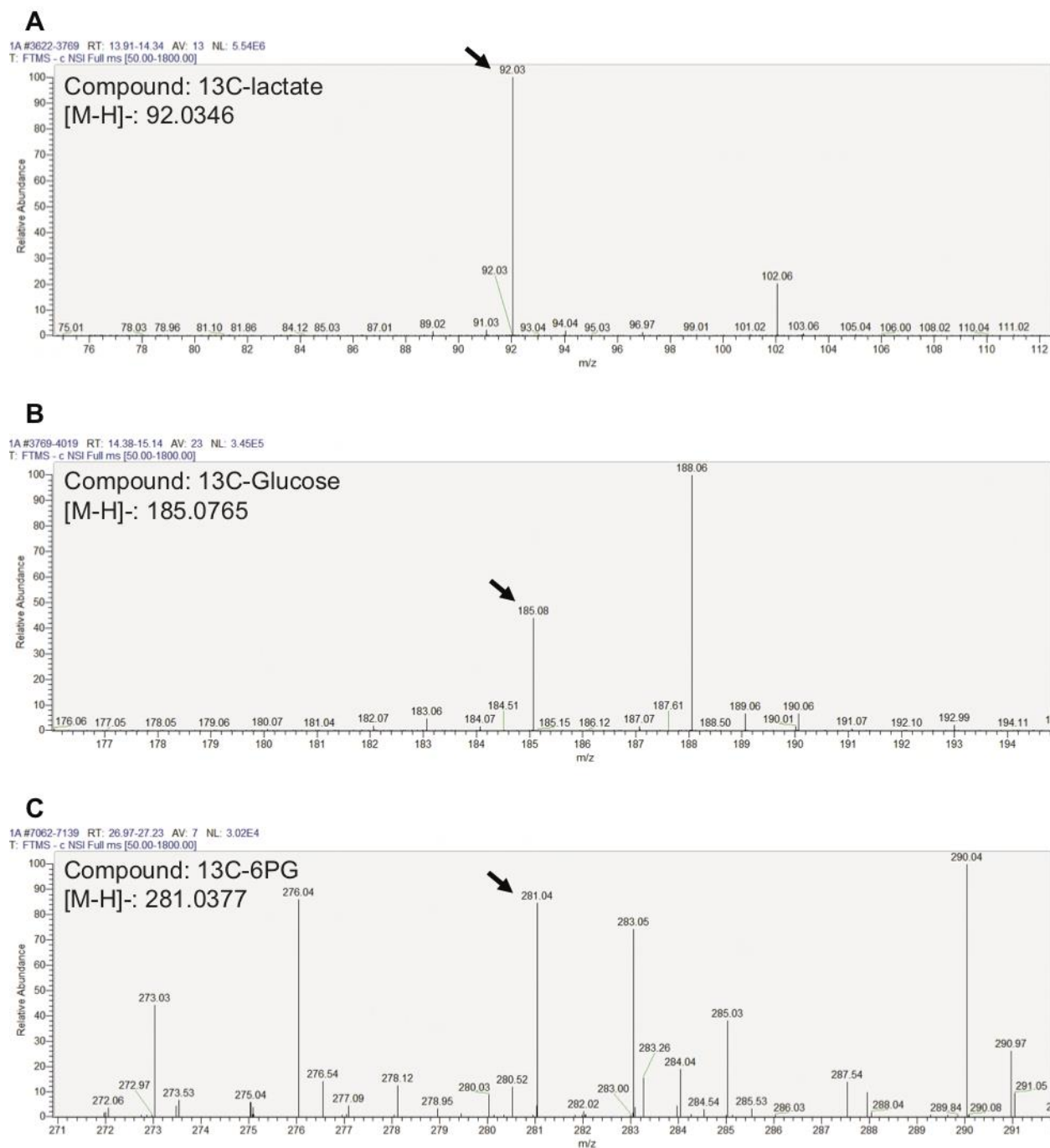


Figure 2.8. Representative mass spectra for ^{13}C -labeled metabolites in a CFME reaction incubated at 37 °C for 1 h. A) lactate, B) glucose, and C) 6-phosphogluconate (6PG).

Table 2.1. List of detected ^{13}C -labeled metabolites, retention times (aligned across samples using MZmine), theoretical fully ^{13}C -labeled negative mode m/z values, m/z values of detected features, and calculated mass errors.

Metabolite	Aligned Retention Time	Max. No. Glucose-derived Carbons	^{13}C labeled m/z (theoretical)	Detected m/z	Mass Error (ppm)
Pyruvate	14.12	3	90.0189691	90.0182423	8.07386736
Lactate	13.07	3	92.0346181	92.0340567	6.09947799
Succinate	14.52	3	120.029534	120.028628	7.55036581
Aspartate	13.30	3	135.040433	135.040497	0.47173434
Acetyl-P	20.88	2	140.986985	140.986189	5.64597728
Histidine	12.86	5	159.079201	159.077028	13.6628837
Glyceraldehyde-3-phosphate/ Dihydroxyacetone phosphate	34.23	3	172.00095	172.00017	4.53856615
Glucose	14.48	6	185.076514	185.075593	4.98060011
Tyrosine	14.20	9	189.097218	189.095695	8.05604078
Ribulose-5-phosphate	19.21	5	234.02888	234.027845	4.42525765
6-phosphogluconolactone	24.08	6	263.027195	263.026148	3.98376788
Glucose-6-phosphate/ Fructose-6-phosphate	14.41	6	265.042845	265.041793	3.97069804
6-phosphogluconate	27.19	6	281.03776	281.036581	4.19826553
Sedoheptulose-7-phosphate	18.44	5	296.056811	296.055299	5.1058181
Fructose-1,6-bisphosphate	24.55	6	345.009176	345.007811	3.95796794
Acetyl-CoA	28.91	2	810.125302	810.123184	2.61456579

CHAPTER 3

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

A version of this chapter was originally published by Jaime Lorenzo N. Dinglasan, David C. Garcia, Him Shrestha, Paul E. Abraham, Robert L. Hettich, and Mitchel J. Doktycz:

Garcia, D.C.,* **Dinglasan, J.L.N.**,* Shrestha, H.K., Abraham, P.E. Hettich, R.L., Doktycz, M.J. (2021). A lysate proteome engineering strategy for enhancing cell-free metabolite production. *Metabolic Engineering Communications*. 12 e00162, doi: 10.1016/j.mec.2021.e00162. *Authors contributed equally

This chapter has been adapted from its published format to accommodate new Figure, Table, and Equation enumeration. The Supplementary Information associated with this work may be found in the Appendix section of this chapter. All references are located at the end of the document.

3.1. 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 biological products without the need to maintain cell viability and enables metabolic engineers to explore novel chemical transformation systems. Crude extracts maintain much of a cell's 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 constrained. Here, we employ Multiplex Automated Genomic Engineering (MAGE) to tag proteins for selective depletion from crude extracts so as to specifically direct chemical production. Specific edits to central metabolism are possible without significantly impacting cell growth. Selective removal of pyruvate degrading enzymes resulted in engineered crude lysates that are capable of up to 40-fold increases in pyruvate production when

compared to the non-engineered extract. The described approach melds the tools of systems and synthetic biology to showcase the effectiveness of cell-free metabolic engineering for applications like bioprototyping and bioproduction.

3.2. 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 potential⁸⁶. 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 metabolites⁸⁷. 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^{57,88}. Constructing complex, multistep pathways require significant development and upfront costs as utilizing purified proteins at scale remains costly⁴⁰. 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 yield⁴⁰. Nevertheless, long running CFME systems that can catalyze multi-step reaction pathways for days have been developed⁵⁷.

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^{46,77,89}. Crude extract systems derived from both commonly used cell-free model organisms, such as *E. coli* BL21 Star (DE3), or nontraditional strains, such as *Vibrio natriegens*, contain a similar biochemistry to the donor cell and can serve as both prototyping tools for *in vivo* metabolic engineering and as bioproduction platforms^{46,90–92}. Cell-free systems work well for both prototyping and production as CFME can be modularly assembled with lysates enriched for specific enzymes or entire metabolic pathways in order to produce a specific molecule⁹³. Additionally, their compatibility with chemical reactors and ability to consume low cost feedstocks have popularized them as potential sources for industrial production^{29,94,95}. These combined capabilities allow CFME processes to make use of tools from traditional bioproduction platforms while taking advantage of the open and modular nature of cell-free systems.

While environmental variables of a cell-free system can be easily 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 (**Figure 3.1**)⁷². 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^{72,96}. Further, specific enzymes can be added or expressed in an extract to further define metabolite production⁸⁶. 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^{46,72}. Tools that allow shaping of

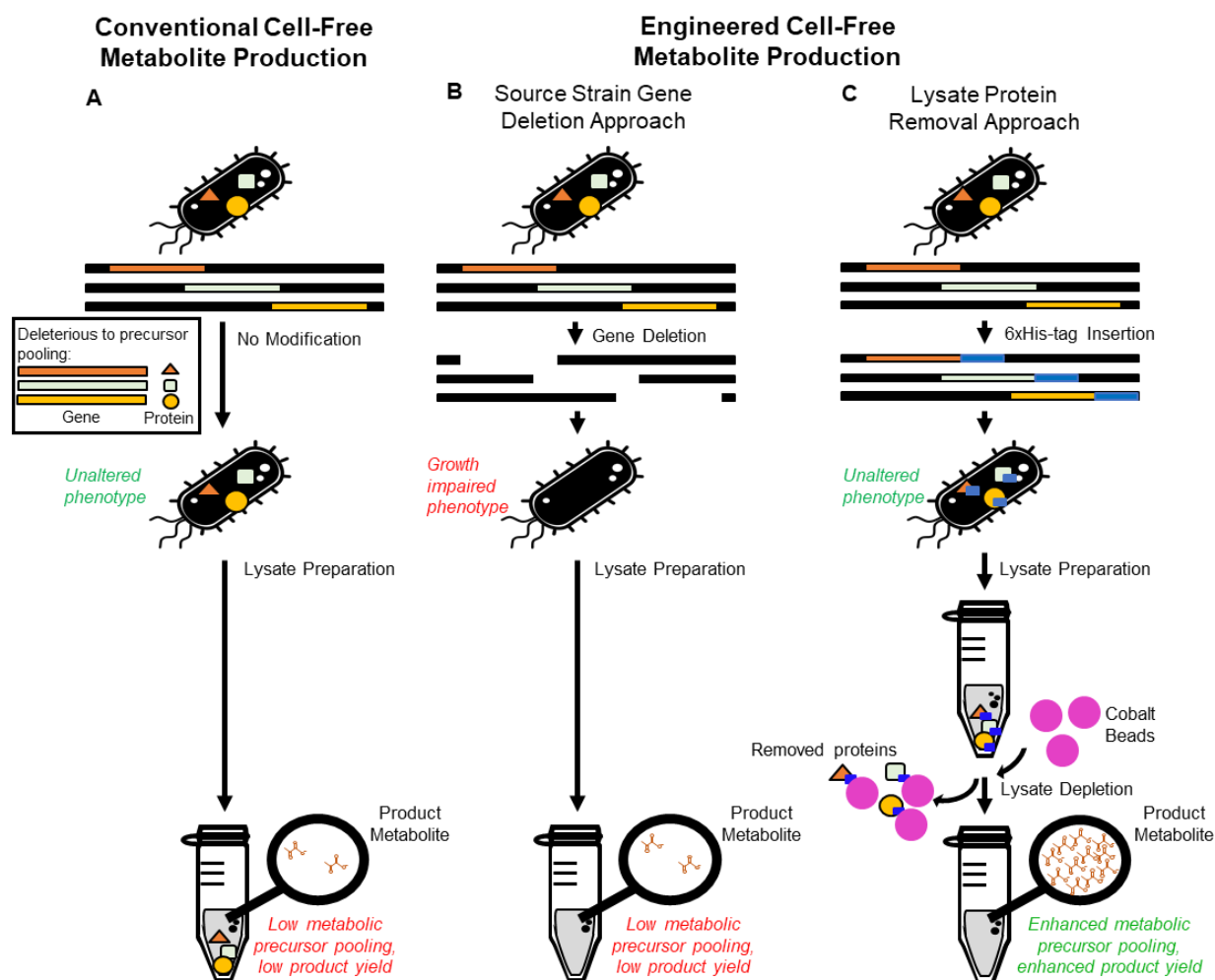


Figure 3.1 Overview of approaches to preparing lysates for cell free metabolic engineering.

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. 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.

the cell-free proteome have been proposed but have not been applied towards the manipulation of cell-free metabolism. Instead, these efforts have focused on improving various single aspects of transcription and translation^{97,98}. Providing approaches with the ability to modulate the presence of multiple enzymes and specific pathways will be critical in enabling the use of crude extract systems for metabolic engineering applications (**Figure 3.1**). 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.

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 depletion 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^{99,100}. 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 3.1**)^{101,102}. This general strategy was demonstrated using *E. coli* BL21 Star (DE3) as a chassis due to its prevalent use for cell-free synthetic biology.

3.3. Methods

3.3.1. 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 (Coralville, IA) with standard purification. Each targeting oligo incorporated four phosphorothioated bases on the 5' terminus. An 18-base CACCATCACCATCACCAT sequence was used to add the 6xHis-tag and directed at either the N- or C-terminus based on previous literature or crystal structure analysis^{103,104}. 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 outgrowth¹⁰⁵. Given the significant time required to find accumulated mutations in a single strain, the additive mutations were started from previously found mutations such that $\Delta 1$ was used to create $\Delta 2$ and so on as per the protocols used in previous studies⁸⁸. After every 8-12 cycles of MAGE, 30-60 colonies were screened for genome edits using MASC-PCR as detailed previously⁸⁸. Allelic genotyping was performed using standard primers designed to flank both modified genes. Amplicons were Sanger sequenced to validate the insertion of the 6xHis-tag sequence. Primer sequences used in this study are listed in Supplemental Table 1 and Supplemental Table 2.

3.3.2 Cell-free extract preparation protocol

Following plasmid curing, the cell extracts were prepared from *E. coli* BL21 Star (DE3) grown at 37 °C in 2xYPT-G (16 g L⁻¹ tryptone, 10 g L⁻¹ yeast extract, 5 g L⁻¹ NaCl, 7 g L⁻¹ KH₂PO₄, 3 g L⁻¹ K₂HPO₄, 18 g L⁻¹ glucose). Cell extracts were prepared by harvesting 50-mL cultures grown in baffled Erlenmeyer flasks to an OD₆₀₀ of 5.0. 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 acetate, 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 gram 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.

3.3.3 Cell-free extract depletions

Cell extracts were depleted for specific proteins 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 μ L 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 and 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 μ L 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. The selective depletions were verified with an anti-6xHis Western Blot.

3.3.4 CFME reaction set-up

Glucose consumption reactions were carried out at 37 °C in 50 μ L volumes using a solution of 100 mM glucose, 18 mM magnesium glutamate, 15 mM ammonium glutamate, 0.2 mM Coenzyme A, 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 same 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 μ L of 5% TCA. The supernatant was centrifuged at 11,000 x g for 5 minutes and directly used for analytical measurements.

3.3.5 Proteomics sample preparation

Samples of both depleted and nondepleted versions of WT, 6xHis-*pflB*, 6xHis-2, 6xHis-3, and 6xHis-4 cell extracts were each prepared in triplicate as follows. Extracts were solubilized in 200 μ L of 4% SDS in 100 mM Tris buffer, pH 8.0. Trichloroacetic acid was added to achieve a concentration of 20% (w/v). 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 μ L of 8 M urea in 100 mM Tris buffer, pH 8.0. Proteins were reduced with 10 mM dithiothreitol incubated for 30 min and alkylated with 30 mM iodoacetamide for 15 min in the dark at room temperature. Proteins were digested with two separate and sequential aliquots of sequencing grade trypsin (Promega) of 1 μ g. Samples were diluted to 4 M urea and digested for 3 hours, 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 manufacturer's instructions. Samples were vacuum-dried with a SpeedVac (Thermo Scientific) and then resuspended in 50 μ L of 0.1% formic acid. Peptide concentrations were then measured using a NanoDrop spectrophotometer (Thermo Scientific) and 2 μ g of each sample was used for LC-MS measurement.

3.3.6 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 μ m inner diameter and 15 cm total length) coupled to an in-house-pulled nanospray emitter of 75 μ m inner diameter packed with 25 cm of 1.7 μ m of Kinetex C18 resin (Phenomenex). For each sample, a single 2 μ g injection of peptides

were loaded and analyzed across a salt cut of ammonium acetate (500 mM) followed by a 210 min split-flow (300 nL/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.

3.3.7 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 false discovery rate 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.

3.3.8 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.

3.3.9 Metabolite measurements

Glucose, pyruvate, lactate, acetate, formate, and ethanol measurements were performed via high-performance liquid chromatography (HPLC) with an Agilent 1260 equipped with an Aminex HPX 87-H column (Bio-Rad, Hercules, CA). 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. Metabolite concentrations were calculated from measurements collected through a refractive index detector (Agilent, Santa Clara, CA) and a diode array UV-visible detector (Agilent, Santa Clara, CA) reading at 191 nm. Pyruvate, glucose, lactate, acetate, formate, and ethanol standards were used for sample quantification using linear interpolation of external standard curves.

3.4. Results and Discussion

3.4.1. Genome engineering enabled display of 6xHis-tags on endogenous pyruvate-consuming enzymes

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^{99,100}. In order to create a pyruvate pooling phenotype in an *E. coli* cell-free extract, four proteins were chosen as targets for removal, LdhA, PflB, AceE, and PpsA (**Figure 3.2**) (**Table 3.1**)¹⁰⁰. These were chosen due to their direct role in consuming pyruvate as well as the likelihood that they are active in lysates. LdhA was selected as the production of lactate from pyruvate has been observed in cell-free systems generated under similar conditions to those prepared here^{71,78}. Since LdhA has previously been reported to be absent in lysates derived from aerobically grown cultures, we assumed that

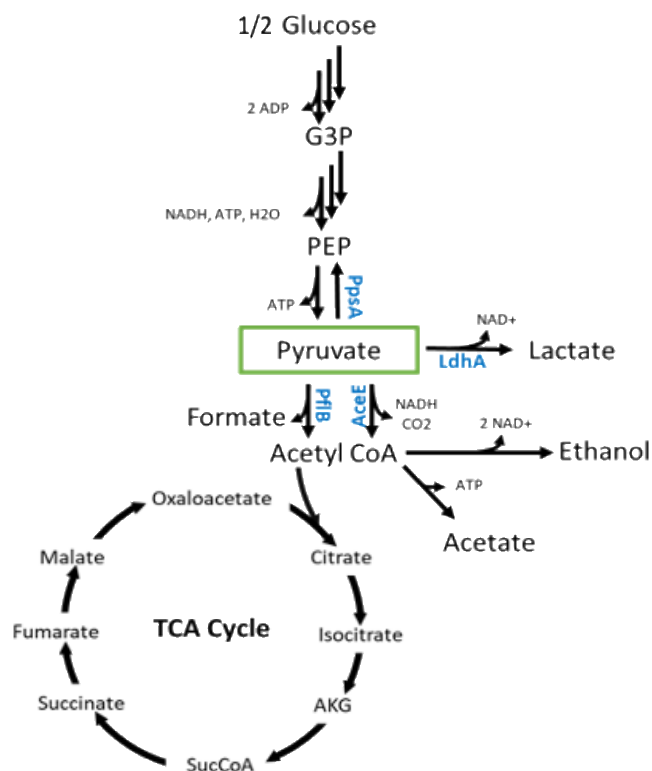


Figure 3.2. Glycolysis and engineered pathway nodes showing the location of the modified enzymes PflB, LdhA, PpsA, AceE.

Table 3.1. Gene and protein information for MAGE targets with a potential effect on pyruvate metabolism.

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

oxygen-limited zones are present in the cultures upon harvesting at mid-log phase¹⁰⁶. This is typical for cells grown in flasks, even under constant shaking, in media with high concentrations of glucose such as the 2xYPTG media (18 g L⁻¹ glucose) used here^{107,108}. Under this assumption, PflB is also likely expressed minimally in the culture and would be capable of carrying pyruvate flux for glycolytic fermentation^{109,110}. At the same time, pyruvate dehydrogenase (Pdh), responsible for pyruvate flux in aerobic conditions, is also expected to be expressed in under these conditions as respiratory metabolism is reportedly active in S30 lysates¹¹¹. Pdh expressed in oxygen-limited cultures is additionally known to be functional in *E. coli* lysates as long as NADH concentrations are not allosterically inhibiting^{112,113}. During a cell-free metabolic reaction, one might expect PflB to be inactivated by oxygen due to its oxygen-sensitivity, and for pyruvate to acetyl-CoA flux to be controlled by Pdh¹¹⁴. Such activity however has yet to be reported so both PflB and AceE, the E1 component of Pdh, were additionally chosen as targets here. We also selected PpsA under the assumption that back-flux to phosphoenolpyruvate (PEP) might occur upon high pyruvate pooling in the lysates (**Figure 3.2**). 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 in order to find a non-inhibitory location (**Table 3.1**)^{103,104}.

The pORTMAGE system was used instead of the traditional genome integrated system due to its potential transferability to multiple donor organisms including *E. coli* BL21 Star(DE3)^{115,116}. Additionally, pORTMAGE is curable following genome engineering and relieves the metabolic burden on the cell that can be imparted due to plasmid maintenance³⁶. Colony screening was performed using MASC-PCR and further verified using Sanger sequencing (**Figure 3.3B**)¹¹⁷. A total of 5 strains were used for this work. (**Table 3.2**). The strains included 6xHis-*pflB*, 6xHis-2 (6xHis-*pflB*-*ldhA*), 6xHis-3 (6xHis-*pflB*-*ldhA*-*ppsA*), 6xHis-4 (6xHis-*pflB*-*ldhA*-*ppsA*-*aceE*) and 6xHis-*ldhA* with each having a varying metabolic phenotype. 60 rounds of MAGE were needed to incorporate all four of the tags into the *E. coli* genome (**Figure 3.3A**) (**Table 3.2**).

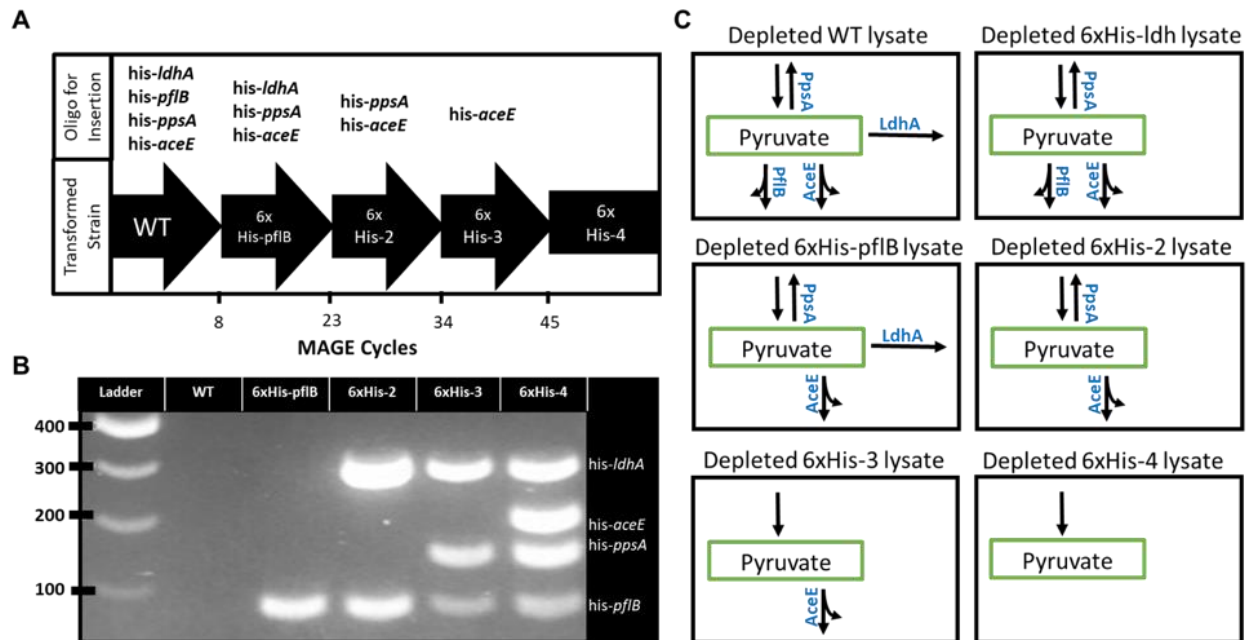


Figure 3.3. Source strain multiplex genome engineering and expected metabolic phenotypes of derived lysates post-depletion. **A)** Strain construction course by MAGE cycling culminating with the 6xHis-4 containing all 4 tags. Each arrow designates the strain being taken through the MAGE process with the oligos used to transform each strain above the arrow. **B)** MAGE-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 depletion of lysates derived from all generated strains.

Table 3.2. Strains created for this study.

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

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 system⁸⁸.

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 3.6**)^{118,119}.

3.4.2. Analysis of glucose metabolism and pyruvate production in enzyme-depleted extracts

The effect of proteome reduction on the extract's metabolic profile was then tested by measurement of glucose consumption, pyruvate accumulation, and the pooling of fermentation end-products (i.e., lactate, ethanol, formate, and acetate) in a CFME reaction mix. As nonspecific binding is commonly associated with the use of 6xHis-tags, we evaluated whether the reduction method would result in significant alterations in lysate metabolism. Evidently, the wild-type derived lysate and the wild-type lysate taken through the depletion process have comparable glucose consumption and fermentation end-product pooling (**Appendix Figure 3.7A**). Initial concentrations of acetate from the S30 buffer used to prepare the extracts could be observed (**Appendix Figure 3.7**). Further, there is no apparent pyruvate accumulation after incubation of the WT lysates with cobalt beads, indicating that the depletion process does not remove proteins that affect cell-free pyruvate production in an appreciable manner (**Appendix Figure 3.7A**). Extracts derived from 6xHis-*pflB*, 6xHis-*ldh*, 6xHis-2, 6xHis-3, and 6xHis-4 were thus reduced and assessed for glucose consumption and pyruvate build-up relative to their unreduced counterparts (**Figures 3.4A & B**). Significantly, none of the nondepleted extracts derived from these strains accumulated pyruvate, and metabolite profiles all trended similarly in terms of

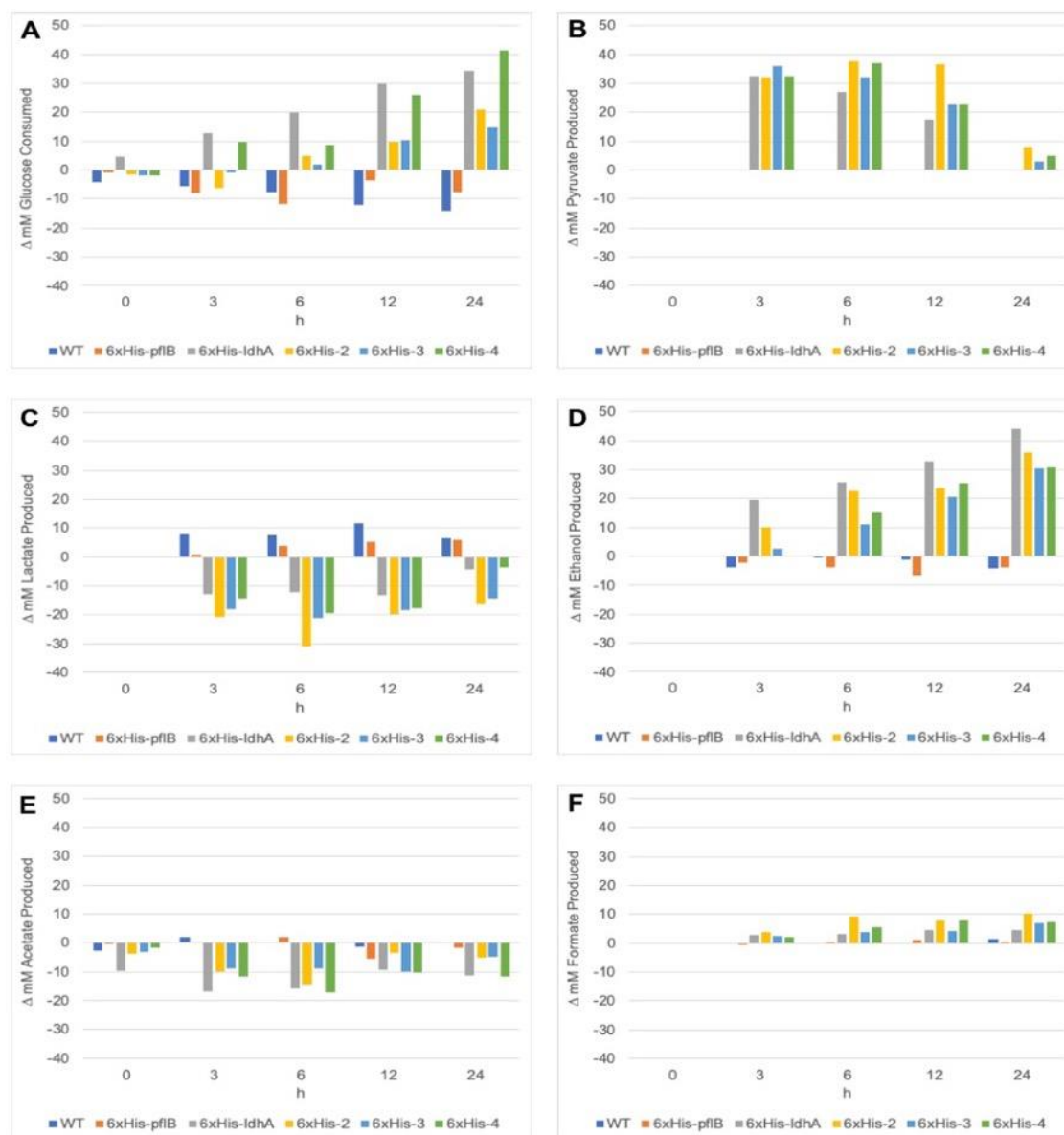


Figure 3.4. Relative changes between nondepleted and depleted versions of the lysates.

A) glucose consumption (nondepleted minus depleted), and **B)** pyruvate, **C)** lactate, **D)** ethanol, **E)** acetate, and **F)** formate production (depleted minus nondepleted) over time. Depleted extracts have had specific 6xHis-tagged proteins removed by incubating with them cobalt beads. Extracts containing tagged proteins, but without an incubation step, are referred to as nondepleted. Data for the time course reactions were acquired using n=3 biological replicates. Standard errors calculated for replicates were negligible and are visualized in **Appendix Figure 3.7**.

glucose consumption and fermentation end-product pooling (**Appendix Figure 3.7**). A noteworthy general observation from the metabolite profiles of the depleted 6xHis-*pflB*, 6xHis-*ldh*, 6xHis-2, 6xHis-3, and 6xHis-4 lysates is that they all continue to accumulate downstream products of the target pyruvate-consuming enzymes, albeit with varying trends (**Appendix Figures 3.7B to F**). This would suggest that the depletion method did not completely remove the targeted enzymes from the lysate proteome, but evidently allows a degree of targeted protein depletion that results in significant metabolic changes.

3.4.3. Analysis of glucose metabolism and pyruvate production in enzyme-depleted extracts

The effect of proteome reduction on the extract's metabolic profile was then tested by measurement of glucose consumption, pyruvate accumulation, and the pooling of fermentation end-products (i.e., lactate, ethanol, formate, and acetate) in a CFME reaction mix. As nonspecific binding is commonly associated with the use of 6xHis-tags, we evaluated whether the reduction method would result in significant alterations in lysate metabolism. Evidently, the wild-type derived lysate and the wild-type lysate taken through the depletion process have comparable glucose consumption and fermentation end-product pooling (**Appendix Figure 3.7A**). Initial concentrations of acetate from the S30 buffer used to prepare the extracts could be observed (**Appendix Figure 3.7**). Further, there is no apparent pyruvate accumulation after incubation of the WT lysates with cobalt beads, indicating that the depletion process does not remove proteins that affect cell-free pyruvate production in an appreciable manner (**Appendix Figure 3.7A**). Extracts derived from 6xHis-*pflB*, 6xHis-*ldh*, 6xHis-2, 6xHis-3, and 6xHis-4 were thus reduced and assessed for glucose consumption and pyruvate build-up relative to their unreduced counterparts (**Figures 3.4A & B**). Significantly, none of the nondepleted extracts derived from these strains accumulated pyruvate, and metabolite profiles all trended similarly in terms of glucose consumption and fermentation end-product pooling (**Appendix Figure 3.7**). A noteworthy

general observation from the metabolite profiles of the depleted 6xHis-*pflB*, 6xHis-*ldh*, 6xHis-2, 6xHis-3, and 6xHis-4 lysates is that they all continue to accumulate downstream products of the target pyruvate-consuming enzymes, albeit with varying trends (**Appendix Figures 3.7B to F**). This would suggest that the depletion method did not completely remove the targeted enzymes from the lysate proteome, but evidently allows a degree of targeted protein depletion that results in significant metabolic changes.

The targeted depletion of PflB from the 6xHis-*pflB* extract results in a metabolic profile that is similar to its control counterpart in that neither accumulate pyruvate (**Figure 3.4B**). Changes in glucose consumption and lactate, ethanol, acetate, and formate production between the depleted 6xHis-*pflB* and WT lysates, relative to their nondepleted controls, are also insignificant (**Figures 3.4A, C to F**). This metabolic phenotype could be expected considering that Pdh activity is active in crude extracts of *E. coli*¹¹². The Pdh complex has a higher affinity for pyruvate than PflB ($K_m = 0.4$ versus 2.0 mM, respectively) when its activity is not inhibited by NADH such as in the presence of an NADH sink like LdhA (**Figure 3.2**)¹⁰⁹. Thus, regardless of whether PflB concentrations are decreased, Pdh could likely be an active route for flux towards ethanol and acetate in this lysate.

The lysate with targeted depletions of both PflB and LdhA (6xHis-2, depleted) pooled 32 mM pyruvate relative to its nondepleted control in 3 h (**Figure 3.4B**). This lysate additionally consumed glucose steadily while maintaining a > 30 mM pyruvate concentration until 12 h. The depletion of LdhA is supported by the observation of a lower lactate concentration in reactions run in these lysates compared to their nondepleted counterparts (**Figure 3.4C**). The rapid build-up of ethanol (20 mM in 3 h relative to the control) in these lysates, and the likely increased activation of aldehyde-alcohol dehydrogenase (AdhE) as an alternative sink for NAD⁺ regeneration, also supports successful LdhA depletion (**Figure 3.2, Figure 3.4D**). Acetate production is also not observed after LdhA depletion, presumably pointing to the increased funneling of acetyl-CoA from the Pdh reaction towards ethanol production to meet redox balance

(**Figure 3.4E**). Rather, acetate seems to be increasingly consumed as a secondary carbon source likely for generating more acetyl-CoA through the acetyl-CoA synthetase route (**Figure 3.4E**)¹²⁰. On the other hand, the contribution of depleted PflB to the observed metabolic phenotype in reactions run in this 6xHis-2 lysate is not as immediately observable. However, at time points before 12 hours, there is a notable decrease in lactate and an increased maintenance of high pyruvate concentrations (**Figures 3.4B & C**). In comparison, the depleted 6xHis-*ldhA* lysate was not as efficient at retaining high pyruvate concentrations as the depleted 6xHis-2 lysate (**Figure 3.4B**). The depletion of both LdhA and PflB from the 6xHis-2 extract may funnel pyruvate flux through Pdh, a claim bolstered by previous work showing NADH-insensitive Pdh to limit glucose fermentation in the absence of both PflB and LdhA¹⁰⁹. Thus, the pyruvate pooled up to 12 h in the depleted 6xHis-2 lysate likely results from lowered glycolytic rates in these extracts. This interpretation is supported by the lowered consumption of glucose in the 6xHis-2 lysate compared to the 6xHis-*ldhA* lysate (**Figure 3.4A**). The concomitant increase in ethanol production and significantly lowered lactate synthesis at 3 h and 6 h in the depleted lysate relative to its control additionally suggests that pyruvate flux is directed through Pdh-AdhE maintaining redox balance by generating net 1 mol NAD⁺ per mol pyruvate (**Figures 3.4C & D**)¹⁰⁹. Compared to reactions run when depleting PflB and LdhA individually, the co-depletion of LdhA and PflB has an additive effect on cell-free pyruvate pooling. This interpretation suggests that oxygen-sensitive PflB is indeed active in *E. coli* crude extracts, which is supported by the observable production of formate after LdhA reductions (**Figure 3.4F**). We reason that decreasing the concentrations of the NAD⁺ regenerating LdhA enzyme limits the *in vitro* activity of formate dehydrogenases that require NAD⁺ as a substrate to decompose formate to CO₂ and H₂O, thus resulting in formate build-up¹²¹. Compared to the depleted 6xHis-2 lysate, the pull-down of PpsA from the 6xHis-3 lysate led to a steady decrease of the pyruvate concentration after 3 h (**Figure 3.4B**). This observation presumably points to the importance of PpsA as an ATP sink in *in vitro* metabolism. *E. coli* glycolytic flux is naturally responsive to the cell's energy charge via the allosteric inhibition of

phosphofructokinase and pyruvate kinase by ATP^{122,123}. The build-up of ATP in the depleted 6xHis-3 lysate that results from glycolysis can lead to lower pyruvate production from glucose at later timepoints, a claim additionally supported by the decrease in relative glucose consumed at 24 h between the 6xHis-2 and 6xHis-3 lysates (**Figure 3.4A**). We additionally observed lower productivities (15 mM in 3 h) and final titers (41 mM) of ethanol formation in the depleted 6xHis-3 extract compared to reactions run in the depleted 6xHis-2 lysate (productivity = 31 mM in 3 h; titer = 55 mM) (**Figure 3.4D, Appendix Figure 3.7C & D**). The low initial accumulation of ethanol despite high pyruvate pooling from LdhA depletion is possibly due to decreased Pdh activity under high adenylate charge¹²⁴. The same condition can explain lowered ethanol production in the depleted 6xHis-3 extract compared to the depleted 6xHis-2 lysate (**Figure 3.4D**). Alternatively, the less efficient ethanol pooling can be due to lowered synthesis rates of NADH from glycolysis after PpsA pull-down.

The targeted depletion of AceE, a component of Pdh, did not increase pyruvate pooling capabilities, but led to the highest consumption of glucose observed (**Figure 3.4A, Appendix Figure 3.7F**). We reason that perturbing the redox balance in the lysate through the pull-down of AceE and thus the depletion of NAD⁺-dependent Pdh activity increased the availability of NAD⁺ for increased glycolytic flux (**Figure 3.2**)¹²⁵. Moreover, the depletion of Pdh activity seems to shift the maintenance of redox balance back to LdhA at later timepoints, as suggested by the steady increase in lactate levels with the decrease in glucose stores (**Figure 3.4A & C**). NAD⁺ is thus continually regenerated by remaining LdhA and ethanol production for the NADH generating step in glycolysis, but this could possibly result in the build-up of glycolytic intermediates since the total consumption of 100 mM glucose is not fully accounted for by the concentrations of pooled fermentation end-products. In general, the rapid consumption of the NAD⁺ supply could be limiting due to the potential lack of cofactor recycling initiated by the decrease of LdhA levels. 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 and further shows that a balance between glucose and pyruvate consumption leads to the engineered pyruvate pooling phenotype (**Figure 3.4B**, **Appendix Figure 3.8**).

3.4.4. Evaluation of modified lysate proteomes after selective enzyme depletion

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 successfully detected (**Figures 3.5B to E**). Although the decrease in PflB levels between the nondepleted and depleted 6xHis-4 lysates met the significance threshold ($p\text{val} < 0.05$), the change was only a 1.81-fold reduction compared to the significant decreases in LdhA and PpsA following lysate depletion (**Figure 3.5E**). AceE was not observed to be pulled down after the purification of 6xHis-tagged proteins from the 6xHis-4 lysate. These findings are inconsistent with metabolic output data as the depletion of 6xHis-4 lysate causes a more significant glucose consumption phenotype than extracts with fewer tags (**Figure 3.4A**). However, anti-6xHis western blots of eluants from the cobalt beads used to deplete the engineered extracts show an enrichment for each of the targeted proteins in their respective strains while no enrichment was seen in the elution from the WT (**Appendix Figure 3.9**). The corroborating evidence of the targeted metabolite analyses combined with antibody tagging leads us to conclude that some level of the targeted proteins are being sufficiently removed and affect the reactivity of the extracts. The minor inconsistency in the results from mass spectrometry and western blot can be explained away as differences between the analytical technique rather than biological. Mass spectrometry is known to identify and quantify proteins with higher reliability, reproducibility, specificity, and sensitivity in complex samples in compared to western blot ^{126,127}. In this case, the comparative proteomic analyses may be complicated by background variability and thus comparability among the nondepleted lysates, depleted lysates, and the eluants ¹²⁸. Although the western blot showed the enrichment result as expected, this result needs to be interpreted with caution.

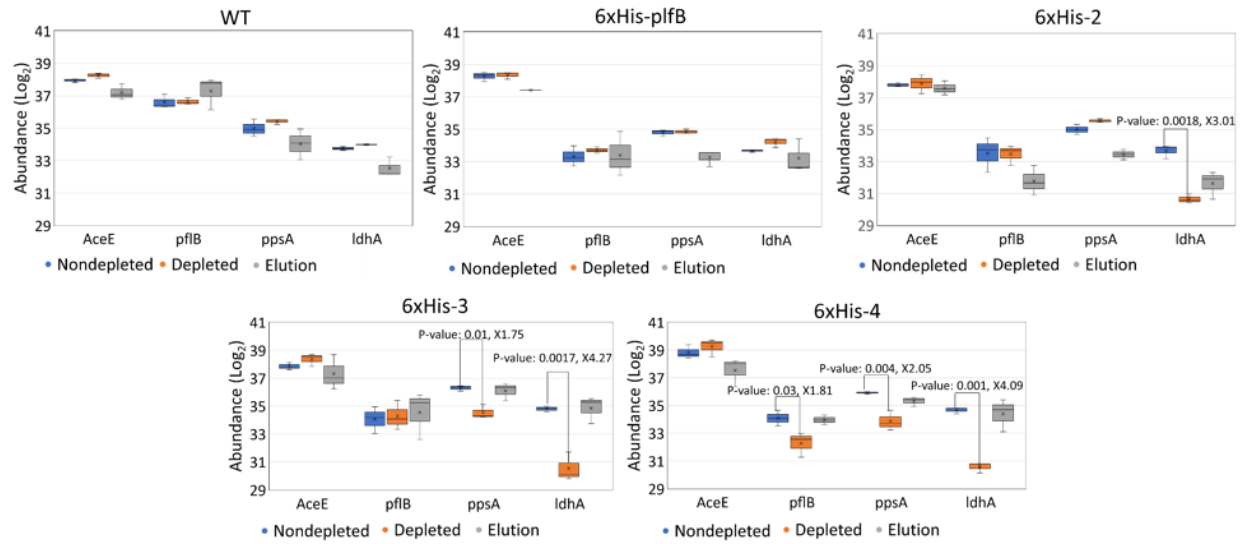


Figure 3.5. Proteomic analysis of control cell-free extracts without depletion (blue), and a depleted cell-free extract (orange), and the elutions from the depleted extracts (gray). Significant fold-changes in protein concentration when comparing the depleted to the nondepleted extract are denoted by p-value and fold change reduction in concentration of the protein above a bracket. Panels show protein levels in the **A)** WT, **B)** 6xHis-pflB, **C)** 6xHis-2, **D)** 6xHis-3, and **E)** 6xHis-4 strains. Each experiment is derived from n=3 biological replicates.

First, the western blot was carried out only in eluant protein fraction whereas the comparative mass spectrometry analysis was carried out in three different protein fractions (i.e., cell extracts with/without targeted depletion and elutions). Thus, the western blot only provides partial information. Second, while bands corresponding to the approximate molecular weights of our targets consistently appeared in eluant fractions where expected, the western blot showed some false positive bands as well (**Appendix Figure 3.9**). Therefore, the results from western blot were utilized as an orthogonal check on the validity of MS results rather than the sole analytical technique for this study. While western blot analysis validated the depletion of proteins not identified through comparative mass spectrometry analysis, future efforts can benefit from a more targeted proteomics approach using labeled peptides to determine absolute quantitative measurements of the method's depletion efficiency for targeted proteins.

Nonetheless, our comparative mass spectrometry analysis provided additional information about our method. The results show that the incorporation of 6xHis-tags into the genomes had minimal effects on the expression of pyruvate-consuming enzymes in all strains' proteomes (**Figure 3.5, blue boxes**), allowing the pull-down of one target protein without altering the concentrations of other pyruvate-consuming enzymes. This observation is corroborated by the comparable trends among the metabolite profiles of all nondepleted lysates (**Appendix Figure 3.7**). This advances our method for precisely generating unconventional metabolic phenotypes that cannot be achieved via gene deletions, since knock-outs of metabolic genes would incite global proteomic and thus metabolic changes in cells¹²⁹. We further analyzed relative proteomic changes in the nondepleted and depleted extracts to determine whether the method resulted in the removal of off-target proteins. Importantly, the process of depleting the proteome did not seem to significantly impact proteins in central metabolism outside of those targeted by the tagging process. When comparing the depleted and nondepleted 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. Future efforts will seek to minimize off-target effects in order to improve the

general applicability of lysate proteome engineering. Though outside the initial scope of this study as the main prospect was to show the use of enzyme depletion as a tool for CFME, targeted proteomics could be an effective tool to connect concentrations of depleted proteins with their resultant metabolic profile.

Targeted depletion 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. The pORTMAGE system offers the potential for extension of this engineering strategy to other, non-model cell-free chassis. Though not all proteins targeted for depletion could be shown to be depleted in substantial quantities through proteomics, our analysis of the metabolic products and western blot analysis shows clear differences between the extracts following each tagging and only following depletion. In contrast with gene knockout strategies that result in global proteomic changes during source strain cultivation, this method allows the selective removal of protein/s from a lysate proteomic background that is similar to the wild type derived extracts, allowing targeted manipulation of lysate proteomes. Thus, although lysates derived from the deletion of a target gene or the post-lysis depletion of its corresponding protein are expected to have different metabolic phenotypes, our CFME approach could be broadly applied to yield metabolic states that are not traditionally possible in living organisms. Future improvements to lysate proteome engineering could make use of multiplex genome engineering methods that are amenable to the insertion of larger tags as MAGE based methods are naturally limited to low-base pair insertions^{116,130,131}. To further advance the depletion of specific proteins in 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-Ion protease system serves this function through a 27 amino acid long peptide and could allow for titration experiments leading to complete degradation of the proteins of interest^{132,133}. A key factor to note stems from MAGE’s limited throughput when making large additions to the genome. Whereas single base changes can be

added with ease, longer tags such as 6xHis tags, are near the edge of feasibility for MAGE tagging.

3.4.5. Evaluating the feasibility of integrating StrepII®-tags into the genome of the model microbe *Vibrio natriegens*

An emerging model microorganism, *V. natriegens*, can take advantage of a MAGE like process termed MUGENT that allows for significantly longer incorporations⁹². This system has higher editing frequencies, reportedly allowing the multiplex integration of 3-4 ~3 kb sequences within a week¹³⁰. *V. natriegens* is also an appealing host for CFME given its short doubling times¹³⁴, which will shorten lysate preparation, and its faster central metabolism than *E. coli*^{135,136}, which may enable more rapid cell-free bioconversion. We evaluated the feasibility of incorporating multiple StrepII®-tags, which are 9 amino acids long, into this microbe's genome. StrepII®-tags have several advantages over 6xHis-tags for endogenous protein display. These include their high affinity for streptavidin-based resins, affording high specificity, and low impact on the physicochemical properties of proteins^{137–139}. To evaluate whether functional StrepII®-tagged endogenous enzymes can be expressed in genome engineered *V. natriegens*, a strain with 11 StrepII®-tagged genomic loci of genes encoding a complete glycolytic pathway was generated and called ST11 (**Appendix Note, Appendix Figure 3.10**). ST11 does not have significantly impaired growth compared to its wild-type (WT) counterpart (**Appendix Figure 3.11A**). Preliminary metabolomics data reveal that the affinity-based StrepII®-tag purification of derived lysates results in a reconstituted enzyme system capable of generating pyruvate and glycolytic intermediates from glucose after a 15-hour reaction period (**Appendix Figure 3.11B**). Preliminary HPLC data show that low but detectable amounts of pyruvate are produced in these systems 15 h after 10 mM glucose supplementation under unoptimized reaction conditions (**Appendix Figure 3.11C**). The low yield can be attributed to multiple issues such as the defined system lacking cofactor regenerating modules, the pathway being cofactor imbalanced, or significant bottlenecks

in the purified pathway due to a non-stoichiometric pull-down of these enzymes. Nonetheless, these data together suggest that the post-lysis strategy can be modified to leverage a microbe that is more amenable to multiplex genome engineering and StrepII®-tags, overcoming some significant limitations identified in the current approach. However, the higher cost of StrepII®-tags than 6xHis-tags and using a donor strain with less study than *E. coli* can be disadvantages in the advancement of the general extract proteome modification approach.

3.5. Conclusions

Shown in this work was the use of genome engineering to create protein modifications that allow for the control of metabolic activity in cellular lysates. This cell-free metabolic engineering strategy allows for the targeted removal of enzymes that can enable the focused production of metabolites from simple precursors using rapidly prepared crude extracts that would otherwise lead to changes in metabolic state and significant growth defects in living cells^{129,140–142}. 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 as well as improving the ethanol titer of the extract. 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 removal of 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 designed to improve its efficiency will be crucial to advancing the scope and adoption of cell-free biological production.

Appendix

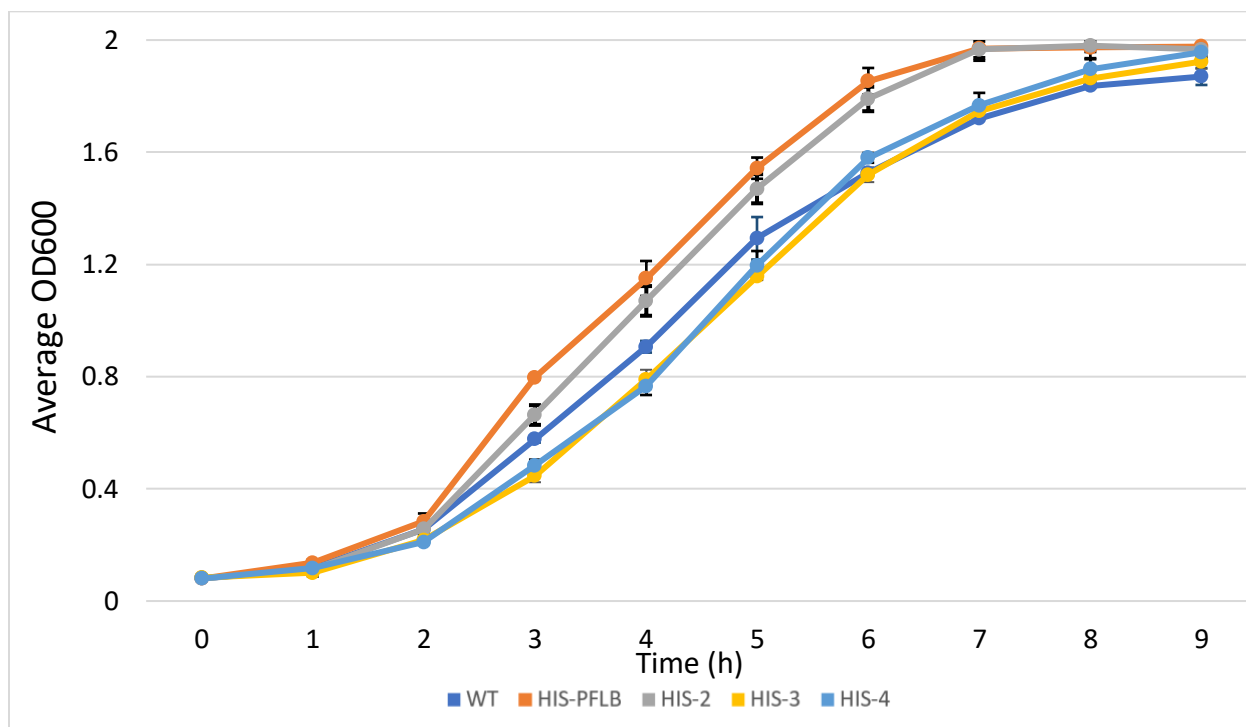


Figure 3.6. Growth rate and terminal OD600 as measured using a WPA CO 8000 Cell Density Meter (Biochrom Ltd., Cambridge, UK) for each of the four main strains generated in this study.

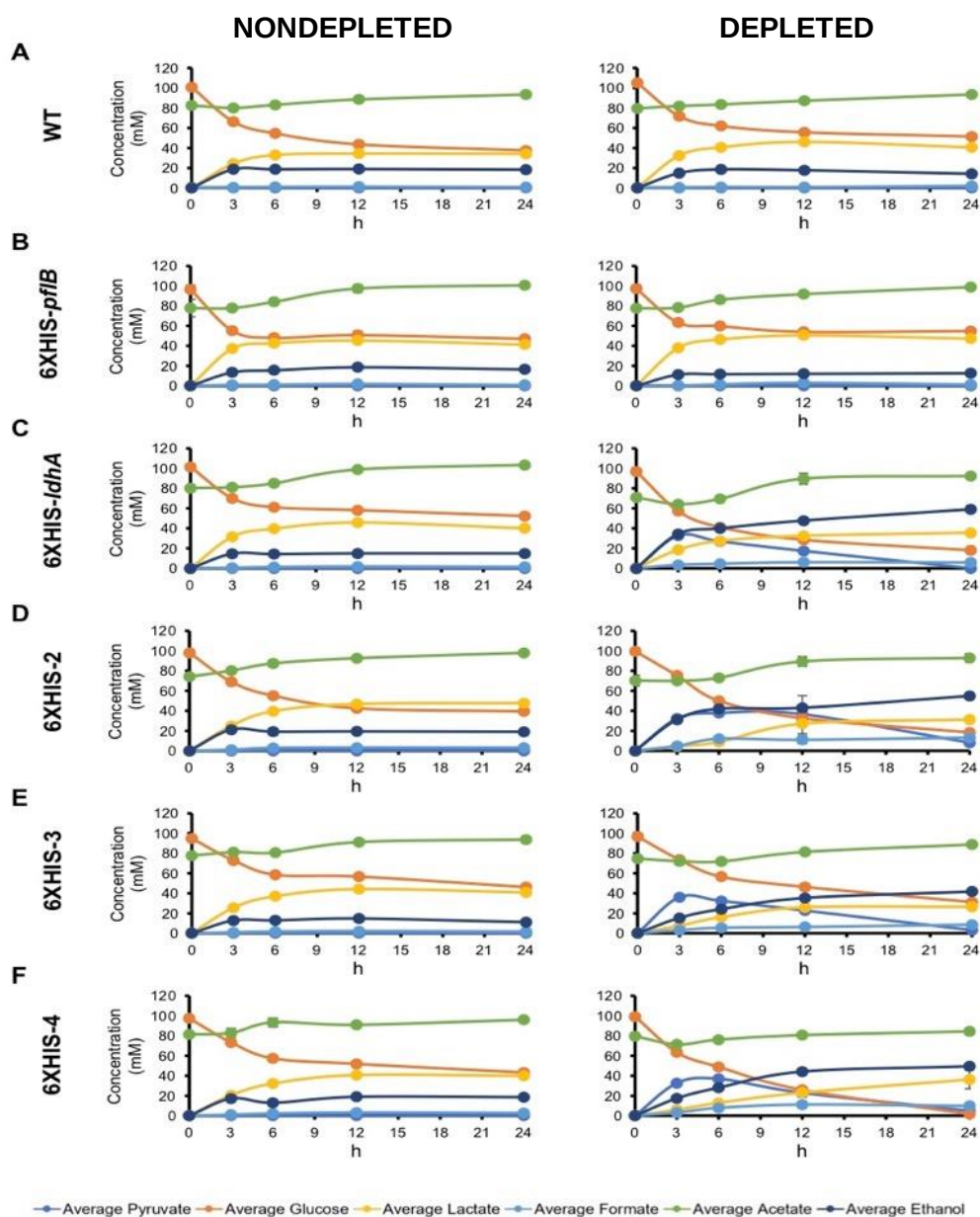


Figure 3.7. Glucose consumption and fermentation product concentrations over time in modified lysates. A) WT, B) 6xHis-pflB, C) 6xHis-ldhA, D) 6xHis-2, E) 6xHis-3, and F) 6xHis-4 lysates. Data and standard deviation for the time course reactions were acquired using $n=3$ biological replicates. Depleted extracts have had specific 6xHis-tagged proteins removed by incubating them with cobalt beads. Extracts containing tagged proteins, but without an incubation step, are referred to as nondepleted.

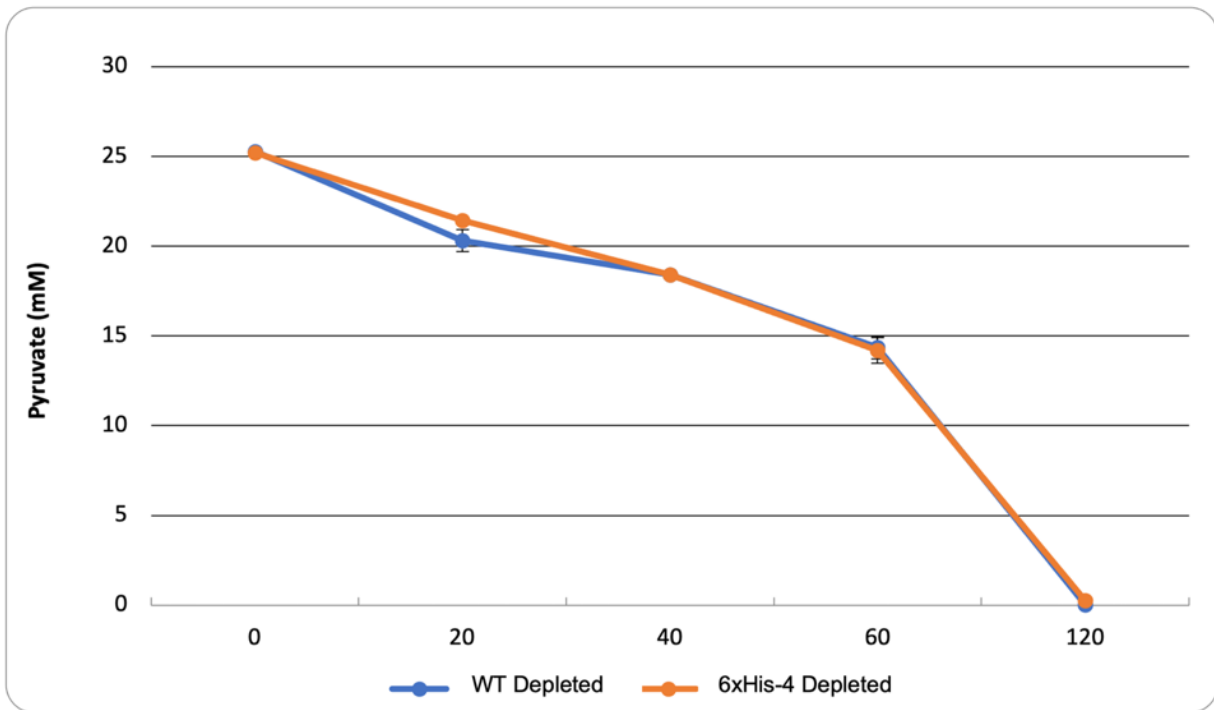


Figure 3.8. Comparison of pyruvate consumption using depleted WT and $\Delta 4$ extracts. All reaction conditions were exactly as those used for the glucose consuming reactions aside from the replacement of glucose with 25 mM pyruvate.

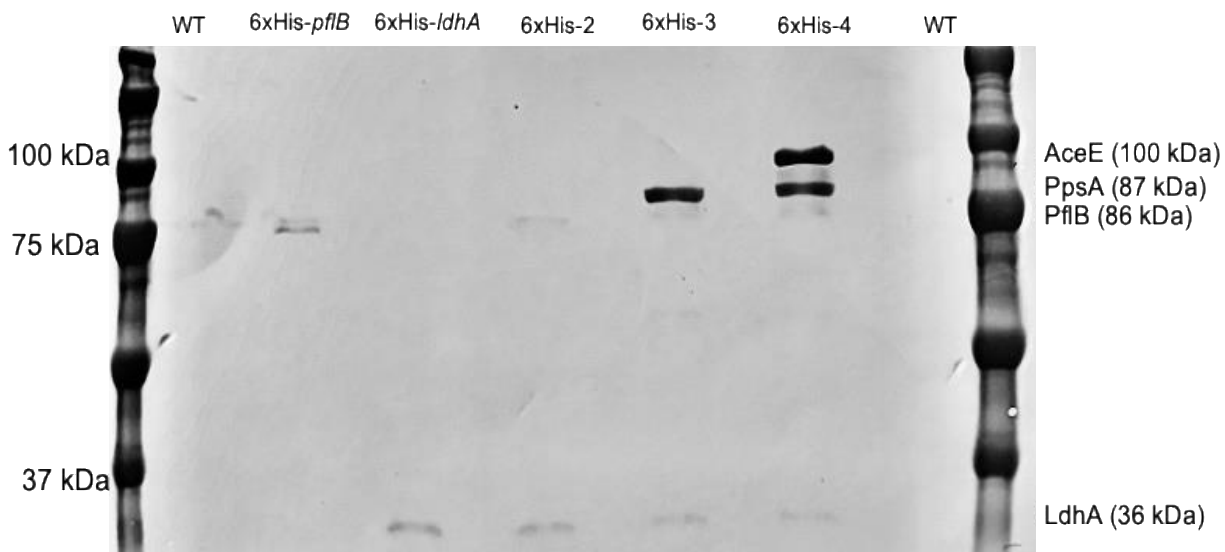


Figure 3.9. Anti-6xHis western blot showing successful pull-downs of target proteins.

Briefly, eluants from cobalt beads used to treat each lysate were concentrated to 0.6 mg/mL total protein concentrations. Protein solutions were denatured with equal volumes of 2x Laemli Sample Buffer (Bio-Rad) and boiled. Samples were loaded onto a 10 % Tris-glycine gel (Bio-RAD) and separated by denaturing PAGE. A Bio-Rad (Hercules, California) Trans-Blot SD Semidry Transfer Cell was used to transfer bands onto a PVDF membrane. The membrane was probed with 6xHis-tag Monoclonal Antibodies (Thermo Fisher Scientific, Waltham, MA) overnight after blocking with Pierce Protein Free (TBS) Blocking Buffer (Thermo Fisher Scientific, Waltham, MA). Blots were incubated with the relevant secondary antibody, HRP conjugate then developed with Pierce DAB Substrate (Thermo Fisher Scientific, Waltham, MA).

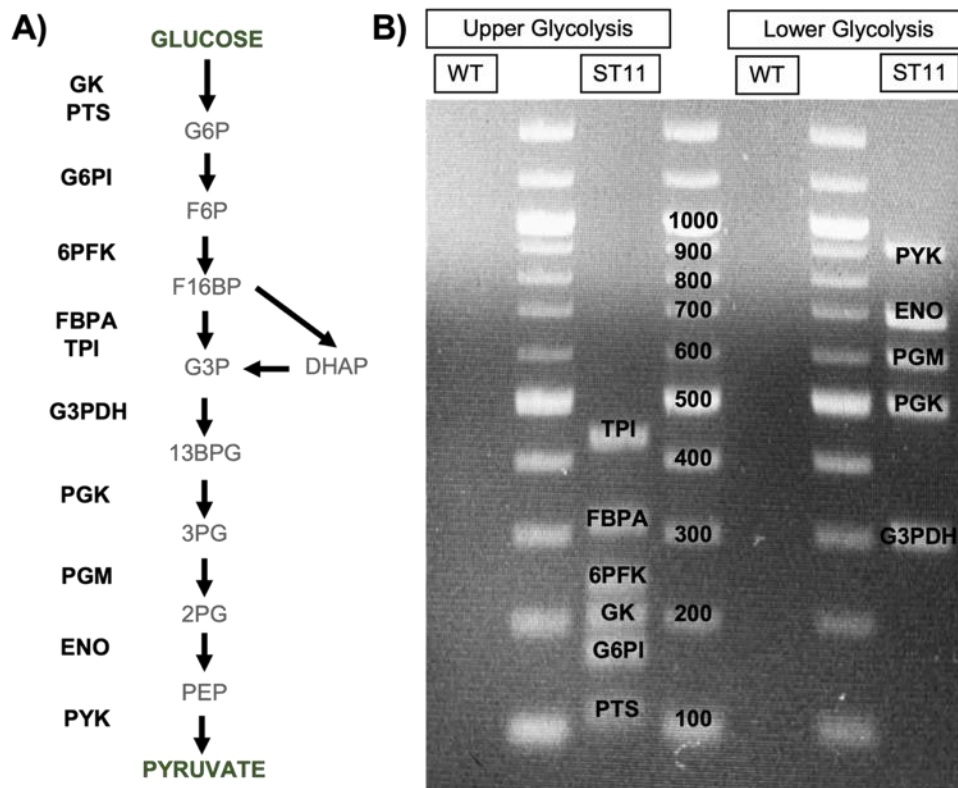


Figure 3.10. StrepII®-tagged glycolytic enzymes in modified *V. natriegens* strain, ST11. A)

Pathway diagram of glycolysis with abbreviations of enzymes targeted for removal in **black bold font**. **B)** Agarose gel of StrepII®-tagged gene sequences in the genome of the ST11 *V. natriegens*

strain, detected via MASC-PCR. MASC-PCR reactions with the same primers were prepared with

the WT *V. natriegens* and no bands were detected. GK = “glucokinase”; PTS =

“phosphotransferase system”; G6PI = “glucose-6-phosphate isomerase”; 6PFK = “6-

phosphofructokinase”; FBPA = “fructose-bisphosphate adolase”; TPI = “triose phosphate

isomerase”; G3PDH = “glyceraldehyde-3-phosphate dehydrogenase”; PGK =

“phosphoglucokinase”; PGM = “phosphoglucomutase”; ENO = “enolase”; PYK = “pyruvate

kinase”; G6P = “glucose-6-phosphate”; F6P = “fructose-6-phosphate”; F16BP = “fructose-1,6-

bisphosphate”; G3P = “glyceraldehyde-3-phosphate”; DHAP = “dihydroxyacetone phosphate”;

13BPG = “1,3-bisphosphoglycerate”; 3PG = “3-phosphoglycerate”; 2PG = “2-phosphogluconate”;

PEP = “phosphoenolpyruvate”.

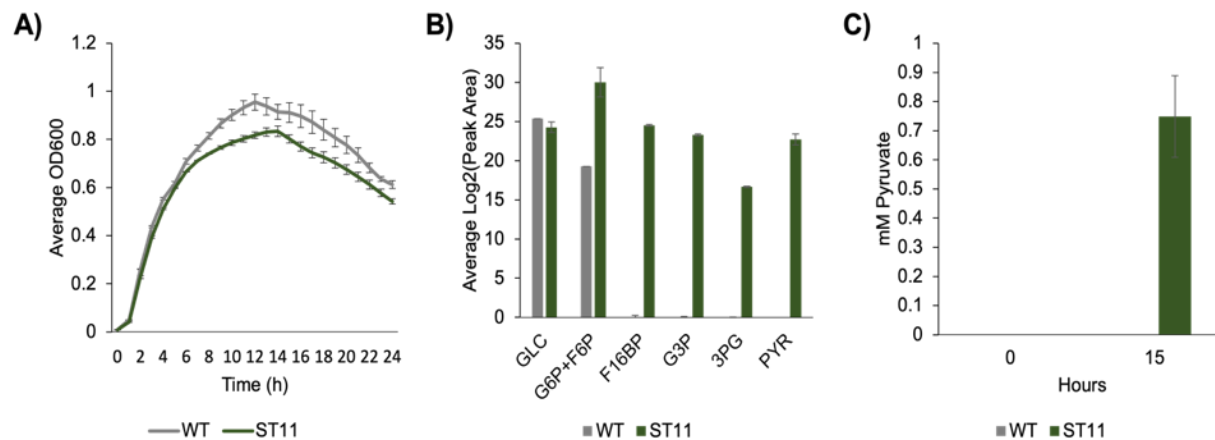


Figure 3.11. Growth curves and metabolite data of strain ST11. **A)** Growth curves of WT and ST11 strains. **B)** LC-MS/MS metabolomics data showing glycolytic intermediates produced in a StreptII®-tag elution of the ST11-derived lysate (n=1) incubated with glucose, cofactors, and salts, and buffer for 15 h (**Appendix Note**). **C)** HPLC-RID quantification of pyruvate in a StreptII®-tag elution of the ST11-derived lysate (n=1) incubated with glucose, cofactors, salts, and buffer for 15 h (**Appendix Note**).

Table 3.3. MAGE oligos use for this study. Phosphorothioated bases are noted with asterisks.

Name	Sequence
Pfl	a*a*t*a*aaaaatccactaagaaggtaggtgttacatgCACcatCACcatCACCATtccgagcttaatgaaa agttagccacagcctgggaa
Ldh	t*a*a*a*tgtgattcaacatcactggagaaagtcttatgCACcatCACcatCACCATaaactcgccgtttatagc acaaaacagtacgacaag
Ppsa	c*a*a*a*ccgttcatttatcacaaaaggattgttcgatgCACcatCACcatCACCATtccaacaatggctcgtc accgctgggtgctttggtat
Pdh	a*c*t*c*aacgttattagatagataaggaataacctatgCACcatCACcatCACCATtcagaacgtttcccaa atgacgtggatccgatcgaa

Table 3.4. MASC-PCR oligos used for this study.

Name	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

Appendix Note:

Protocol for the generation of *V. natriegens* ST11 and the preparation of reconstituted enzyme reactions:

V. natriegens ATCC 14048 was transformed with the pMMB-tfox plasmid and MUGENT was carried out as previously described¹³⁰. Oligos for insertion into *V. natriegens* genomes were prepared by overlap PCR. Briefly, primers were designed so that ~ 3 kb long oligos containing the gene-of-interest and at least 1 kb of flanking sequences upstream and downstream of the gene could be isolated from *V. natriegens* ATCC 14048 genomic DNA. MASC-PCR primers were designed to validate genomic insertions via colony MASC-PCR. After confirming the 11 StrepII®-tagged genes, the ST11 strain was grown to OD 6.0 and harvested for lysis. Lysates were prepared following the same protocol for *E. coli* described here. 500 µL lysate was used to load Strep-Tactin®XT Spin Columns (IBA Life Sciences, #2-4150-025). Columns were washed 4x and eluted for maximum protein yield following the manufacturer's protocol. Approximately 450 µL of elution was buffer exchanged with S30 buffer using Amicon®Ultra Centrifugal Filters 3000 NMWL (Millipore, UFC500324) then concentrated to ~ 90 µL solution.

CFME reactions with these reconstituted enzyme systems were prepared by mixing 35 µL concentrated protein solution with 10 mM glucose, 1.5 mM ATP, 1.5 mM NAD⁺, 10 mM bis-tris buffer, 10 mM dipotassium phosphate, 30 mM tris-acetate, 42 mM magnesium acetate, 180 mM potassium acetate, 18 mM magnesium glutamate, 15 mM ammonium glutamate, 0.2 mM Coenzyme A, and 195 mM potassium glutamate. Reactions were incubated at 37 °C then analyzed via LC-MS/MS or HPLC as described in Chapter 2 of this dissertation.

CHAPTER 4

REWIRING CELL-FREE METABOLIC FLUX IN E. COLI LYSATES USING A BLOCK-PUSH-PULL APPROACH

A version of this chapter has been submitted for publication by Jaime Lorenzo N. Dinglasan and Mitchel J. Doktycz:

Dinglasan, J.L.N. & Doktycz, M.J. (2023). Rewiring Cell-free Metabolic Flux in *Escherichia coli* Lysates using a Block-Push-Pull Approach. *Accepted in Oxford Academic Synthetic Biology*.

This chapter has been adapted from its published format to accommodate new Figure, Table, and Equation enumeration. The Supplementary Information associated with this work may be found in the Appendix section of this chapter. All references are located at the end of the document.

4.1. Abstract

Cell-free systems can expedite the design and implementation of biomanufacturing processes by bypassing troublesome requirements associated with the use of live cells. In particular, the lack of survival objectives and the open nature of cell-free reactions afford engineering approaches that allow purposeful direction of metabolic flux. The use of lysate-based systems to produce desired small molecules can result in competitive titers and productivities when compared against their cell-based counterparts. However, pathway crosstalk within endogenous lysate metabolism can compromise conversion yields by diverting carbon flow away from desired products. Here, the “block-push-pull” concept of conventional cell-based metabolic engineering was adapted to develop a cell-free approach that efficiently directs carbon flow in lysates from glucose and towards endogenous ethanol synthesis. The approach is readily adaptable, relatively rapid, and allows manipulation of central metabolism in cell extracts. In implementing this approach, a block strategy is first optimized, enabling selective enzyme removal

from the lysate to the point of eliminating byproduct forming activity while channeling flux through the target pathway. This is complemented with cell-free metabolic engineering (CFME) methods that manipulate the lysate proteome and reaction environment to push through bottlenecks and pull flux towards ethanol. The approach incorporating these block, push, and pull strategies maximized glucose to ethanol conversion in an *E. coli* lysate that initially had low ethanologenic potential. A 10-fold improvement in percent yield is demonstrated. To our knowledge, this is the first report of successfully rewiring lysate carbon flux without source strain optimization and to completely transform consumed input substrate to a desired output product in a lysate-based, cell-free system.

4.2. Introduction

Cell-based manufacturing systems routinely produce fuels and pharmaceutical precursors and products, among other commercially valuable compounds, from readily available substrates^{13,143,144}. These biomanufacturing systems are commonly evaluated by their productivity (rate of product formation per unit volume), titer (concentration), and yield (amount of product formed per substrate amount consumed)^{13,143,144}. However, engineering a living organism's complex dynamic metabolism while meeting its survival demands typically requires multiple time-consuming design-build-test-learn (DBTL) cycles to make modest gains in these performance metrics^{18,145}. For example, ethanol is a major biofuel and realizing production at high titer (> 40 g/L), productivity (> 1 g/L/h), and yield (> 90% of the maximum theoretical amount of ethanol that can be formed from substrate) in cell-based systems has taken decades of strain optimization^{20,146}. Cell-free optimizations and implementations shows promise in expediting the generation of biosynthetic platforms with high production metrics^{20,57}.

Cell-free or *in vitro* systems employ the components necessary to carry out metabolism outside of a living cell, or in a "cell-free" space^{29,30,145}. These systems are commonly generated

either as reconstituted pathways from purified enzymes (i.e., reconstituted enzyme systems), or as crude cell extracts (i.e., lysate-based systems) via lysis procedures that preserve the functionalities of select cellular components^{18,145}. In either case, cell-free systems are open biochemical environments devoid of organismal survival demands. Their amenability to direct manipulation offers unique and straightforward metabolic engineering approaches and are considered to have high potential for commercially viable production of diverse materials²⁰. For example, lysates have been used to synthesize several commodity chemicals such as 2,3-butanediol and styrene at high titers and productivities^{46,51}. Additionally, cell-free systems can be lyophilized and stored at room temperature which would eliminate the need for a cold chain and enable distributed biomanufacturing⁴¹. This is empowered by ongoing efforts to improve the shelf-stability of these platforms for on-demand small molecule synthesis especially in times of supply chain disruption⁴³.

Implementing cell-free systems requires overcoming key barriers. When rebuilding large conversion pathways, reconstituted enzyme systems can be constrained by high enzyme production and purification costs as well as problems associated with cofactor turnover^{20,46}. On the other hand, lysates typically contain endogenous central metabolic pathways and cofactor regenerating reactions that feed into reactions catalyzed by supplemented heterologous enzymes. Consequently, protein purification requirements are greatly reduced albeit at the expense of reaction crosstalk. However, a significant barrier to production in lysate-based systems is the competition for flux between unwanted endogenous background reactions and target production modules, which inevitably impacts their yield⁴⁷. Making high yield targeted conversions would require insulating an endogenous pathway by removing only a few enzymes that redirect flux through undesired reactions^{20,47,147}. The ability to do so could benefit the rapid cell-free synthesis of high-value metabolites, which commonly have precursors in central metabolism, without arduous whole pathway purification^{19,99}. To advance the commercial potential of extracts for large-scale or distributed biomanufacturing, methods to eliminate unwanted

pathway crosstalk and rewire flux through target endogenous metabolic pathways in lysates are needed.

“Block-push-pull” approaches are commonly applied when engineering cellular metabolism for insulating target pathways and maximizing conversion yields^{148–151}. In this convention, to “block” means to inactivate reactions by removing endogenously expressed enzymes (commonly by gene deletions) that interface with target and flux-competing pathways. To “push” means to drive flux forward by overexpressing flux limiting enzymes, and to “pull” is to direct flux towards the end-product by overexpressing enzyme(s) in a terminal pathway module^{150,151}. These cell-based metabolic engineering strategies however take a toll on host vitality and require laborious molecular cloning procedures, lengthening and increasing DBTL cycles¹⁴⁵. The open nature of lysate-based cell-free systems offers unique push and pull strategies to redirect flux without imposing metabolic burden on source strains. For example, bottlenecks can be overcome and specific pathway modules can be enriched by tailoring the reaction environment (i.e., cofactor availability and pH) to enhance enzyme activity, manipulating source strain cultivation conditions, and supplementing lysates with overexpressed proteins^{17,19,73,78}. On the other hand, reaction blocking strategies that effectively shave lysate endogenous metabolic complexity while taking advantage of the manipulable cell-free environment are lacking. Prior CFME efforts that focused on rewiring carbon flux typically involved deriving lysates from *knock-out* strains^{47,53}. However, this strategy will not always result in predicted changes to lysate metabolism, presumably because gene deletions lead to protein expression changes that rearrange flux distributions and metabolic networks in the source strain¹⁵². This method is also constrained by cell viability and thus faces limitations common to cell-based metabolic engineering⁴⁷. Recently, negative CRISPR effectors that down-regulate byproduct forming enzymes in source strains were shown to rewire metabolic flux in a cell-free system towards target end-products without affecting source strain growth⁶². While this improved product titers, substrate to product conversion yields remained low because this strategy can only

knock-down, and not completely eliminate, flux-competing reactions⁶². Inhibiting enzymes in cell-free reactions may be a viable strategy for inactivating unwanted reactions without impacting source strain phenotypes. This approach also benefits from being able to add inhibitors directly into the open lysate environment. However, the cost and availability of these molecules could be restrictive¹⁵³. Another recently described strategy for reducing concentrations of select enzymes in extracts involves the incorporation of an affinity tag at the N- or C-terminus of endogenously expressed enzymes for subsequent removal from the lysate⁷⁴. Importantly, genomic integration of the 6xHis-tags had no significant effect on source strain viability even though central metabolic, pyruvate-consuming enzymes were targeted. When the derived lysates were fed glucose, the formation of unwanted byproducts was slowed, enabling significant pyruvate accumulation for a few hours followed by eventual byproduct pooling. This affinity-based selective enzyme depletion method can theoretically be improved to completely remove these unwanted reactions post-lysis, but this has yet to be reported.

Here, we describe a cell-free block-push-pull approach that insulates and drives flux through an endogenous lysate metabolic pathway for high-yield metabolic conversion. This is demonstrated with ethanol synthesis from glucose using an *E. coli* derived extract. To achieve this, the affinity-based selective enzyme depletion strategy was optimized to block unwanted reactions, those that compete with flux towards ethanol synthesis. This optimization was performed in lysates derived from a previously described *E. coli* strain that encodes an N-terminal 6xHis-tag at its pyruvate formate lyase B (*pflB*) and lactate dehydrogenase A (*ldhA*) genes. The reaction blocking strategy was then complemented with existing CFME strategies (i.e., reaction and source strain cultivation condition optimizations) that drive flux through bottlenecks and towards ethanol without further genetic manipulations (**Figure 4.1**). Removing unwanted reactions and modifying reaction or cultivation parameters to focus and increase flux through the pathway successfully modified a low ethanol producing *E. coli* lysate into a homoethanogenic cell-free system that produces ethanol from glucose with > 90% yield.

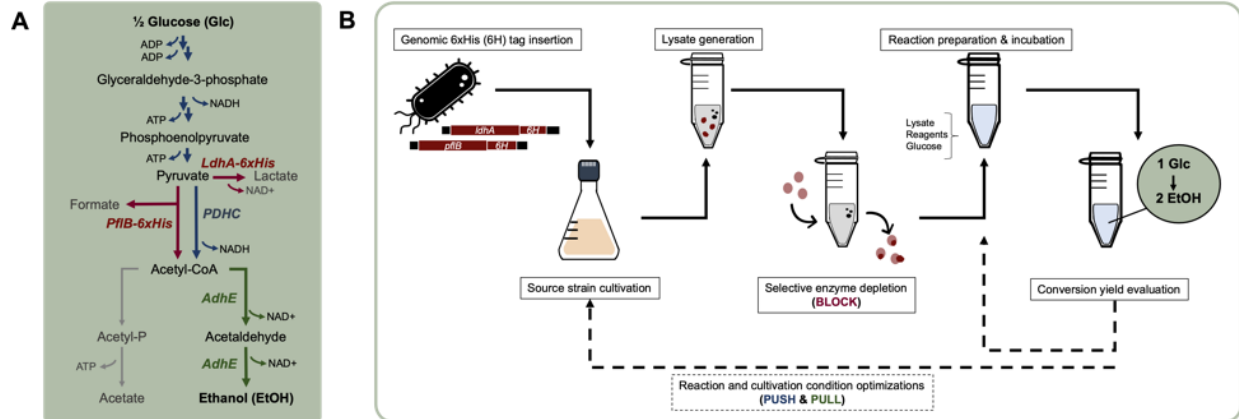


Figure 4.1. Target pathway and workflow for flux optimization. A) Simplified metabolic pathway for endogenous ethanol biosynthesis from glucose in *E. coli* lysates. The strain used here was genome engineered to express endogenous LdhA and PflB proteins with terminal 6xHis tags for selective removal from the derived lysate. Red arrows represent reactions that are targeted for blocking while blue and green arrows represent reactions that flux will be pushed and pulled through, respectively. **B)** Schematic of the cell-free block-push-pull approach employed here to maximize ethanol yield from glucose.

4.3. Methods

4.3.1. Cell Culture

The strain, derived from *E. coli* BL21 (Star) DE3, used here was previously engineered via a pORTMAGE protocol to incorporate genomic 6xHis tag DNA sequence insertions upstream of the *ldhA* and *pflB* genes⁷⁴. These cells were initially grown in standard 2xYPT-G media (16 g L⁻¹ tryptone, 10 g L⁻¹ yeast extract, 5 g L⁻¹ NaCl, 7 g L⁻¹ KH₂PO₄, 3 g L⁻¹ K₂HPO₄, 18 g L⁻¹ glucose). Glucose concentrations in media were adjusted to 9 g L⁻¹ or 27 g L⁻¹ according to our source strain cultivation conditions optimization protocol. Starter cultures in 6 mL 2xYPTG were incubated overnight at 30°C with shaking at 250 rpm. Baffled Erlenmeyer flasks containing 100 mL of the same media were inoculated with 2% of the overnight culture and incubated at 37°C with constant shaking at 250 rpm. Cells were pelleted by centrifugation for 5 min at 10,000 g at 4°C, initially between OD₆₀₀ 5.8-6.0. When optimizing this parameter, cells were strictly grown to either OD₆₀₀ 5.0 or 7.0, then between OD₆₀₀ 5.0-5.2 for subsequent experiments. Cell pellets were washed twice with cold S30 buffer without dithiothreitol (DTT) (14 mM magnesium acetate, 60 mM potassium acetate, and 10 mM Tris-acetate, pH 8.2) then weighed. These were flash-frozen and stored at -80°C unless lysed on the same day. Fresh cultures were prepared for each optimization step.

4.3.2. Lysate Preparation

Cell pellets were resuspended in 0.8 mL S30 buffer (without DTT) per gram wet cell mass, after thawing on ice. Resuspensions were sonicated using a Branson 450 sonifier with a 3.2 mm tip. Briefly, 1 mL resuspension was placed in 1.5 mL microcentrifuge tubes surrounded by an ice water bath. Cells were lysed by applying 590 J at 50% tip amplitude at 10 s ON, 10 s OFF intervals directly to the resuspension. The resulting extracts were centrifuged twice at 21,100 g for 10 min

at 4°C. The supernatant was aliquoted to clean microcentrifuge tubes, flash-frozen, then stored at -80°C until use. Extracts were prepared from fresh cultures for each optimization step.

4.3.3. Lysate Depletion

For optimizing the post-lysis depletion method, HisPur™ Cobalt Resin (ThermoFisher Scientific) suspensions were washed with 600 µL S30 buffer (without DTT) on ice and centrifuged for 30 s at 21,100 g at 4°C. It should be noted that the resin is suspended in ethanol in storage. In initial experiments, all resin suspension volumes were washed twice following our previously reported protocol, but this was insufficient for higher volume resin suspensions as residual ethanol storage buffer affected our time zero sample measurements⁷⁴ (**Appendix Figure 4.8**). More wash cycles remedied this problem (**Appendix Table 1**). The number of wash steps were thus optimized per resin suspension volume to ensure the removal of ethanol buffer used for resin storage. LdhA-6xHis and PflB-6xHis were depleted from lysate proteomes by adding 1X volume of the cell extract to Co²⁺ beads washed from 0.2X, 0.6X, 1X, 1.4X, and 1.8X volume of ice-cold resin suspension. A fixed lysate volume of 200 µL was used. Lysates were initially mixed with beads for 1 hr at 4°C under shaking conditions (900 rpm). Incubation time was changed to 30 min after optimization of this protocol for subsequent experiments. All lysate-bead mixtures were centrifuged at 21,100 g for 30 s. Supernatants were aliquoted, flash-frozen, and stored at -80 °C until use. The depletion method was performed in triplicate on the same batch of extract. Replicates are referred to here as “depletion replicates”.

4.3.4. CFME Reaction Preparation

All cell-free metabolic reactions were prepared with 18 mM magnesium glutamate, 15 mM ammonium glutamate, 195 mM potassium glutamate, 150 mM Bis-Tris, and 10 mM dipotassium phosphate. Lysate was added to these mixtures at a final concentration of 4.5 mg/mL unless otherwise noted. For every tested condition, reactions were prepared with depletion replicate

lysates (n=3) to evaluate variability in the depletion strategy. Cofactor (NAD⁺, CoA, ATP) and glucose concentrations and pH levels used to initiate the reactions are detailed in the manuscript. To achieve acidic pH levels, glacial acetic acid was added to S30 buffer solutions that were then used to prepare the reactions. The pH of the reaction itself was measured with Hydryon MicroFine pH papers prior to incubation. Reactions were carried out in 30 μ L volumes in enclosed 1.5 mL microcentrifuge tubes (Thermo Scientific #3451), 0.2 mL PCR tubes (VWR #20170-012), or 50 μ L wells on a 384-well-plate (Corning #3544) sealed with polyester film (VWR #75853-868) as described. Reactions were incubated at 37°C for 20 h unless otherwise stated. For each replicate, a control time zero reaction was prepared as described previously¹⁵⁴. Briefly, lysate metabolic activity was terminated with 5% TCA prior to the addition of glucose to prepare controls.

4.3.5. Metabolite Analyses

CFME reactions were quenched with an equivalent volume of 5% TCA and then diluted with an equal volume of ultrapure water. High-performance liquid chromatography (HPLC) was used to measure changes in glucose, ethanol, lactate, and acetate concentrations between time zero and endpoint reactions. Measurements were taken with an Agilent 1260 Infinity system (Agilent, Santa Clara, CA) equipped with an Aminex HPX 87-H column (Bio-Rad, Hercules, CA) for carbohydrate, organic acid, and alcohol analyses. An isocratic method using 5 mM sulfuric acid was used to elute target metabolites from the column at a 0.55 mL min⁻¹ flow rate for 30 min at 35°C. Metabolites were detected with a refractive index detector (Agilent, Santa Clara, CA). Peak areas were extracted using manual integration for glucose peaks and automatic integration for lactate, acetate, and ethanol peaks in the HPLC system's companion software. Concentrations were calculated using standard calibration. Reported values for metabolite consumption and production are changes in concentrations between endpoint and time zero reactions. Student's t-tests were used to evaluate statistical significance (pval < 0.05). Percent ethanol yield from

glucose was calculated as follows, given that 1 mole glucose can theoretically be converted to 2 moles ethanol in *E. coli* lysates:

$$\% \text{Yield}_{\left(\frac{\text{EtOH}}{\text{Glc}}\right)} = \frac{\text{Actual Yield}_{\left(\frac{\text{EtOH}}{\text{Glc}}\right)}}{\text{Theoretical Yield}_{\left(\frac{\text{EtOH}}{\text{Glc}}\right)}} \times 100 \quad (1)$$

$$\text{Actual yield}_{\left(\frac{\text{EtOH}}{\text{Glc}}\right)} = \frac{\text{moles EtOH produced}}{\text{moles Glc consumed}} \quad (2)$$

$$\text{Theoretical yield}_{\left(\frac{\text{EtOH}}{\text{Glc}}\right)} = \frac{2 \text{ moles EtOH}}{1 \text{ mole Glc}} \quad (3)$$

4.4. Results and Discussion

4.4.1. Blocking flux-competing pathways via selective depletion of LdhA and PflB proteins in lysates redirects carbon flow to ethanol.

Glucose derived carbon flow could theoretically be diverted towards ethanol by removing PflB and LdhA and associated flux towards formate and lactate. However, the $[\text{NAD}^+]$ to $[\text{NADH}]$ ratio must be balanced to sustain flux towards ethanol. Conventionally, per mole glucose consumed, glycolysis generates 2 moles NADH that are primarily recycled to NAD^+ via the lactate forming LdhA enzyme (**Figure 4.1A**). Because this reaction sustains glycolysis by providing NAD^+ , lactate has been observed as a major fermentation product in lysate metabolism of glucose.¹⁵⁴ To redirect flux towards ethanol, an alternative ethanol forming, $[\text{NAD}^+]/[\text{NADH}]$ balancing pathway can be activated through the PDHC-AdhE pathway module^{74,155}. Briefly, PDHC generates 2 moles acetyl-CoA and 2 moles NADH from 2 moles pyruvate (**Figure 4.1A**). The net output of glycolysis and PDHC is thus 4 moles NADH per mole glucose. AdhE can recycle the

equivalent amount of NAD^+ from 4 moles NADH when transforming 2 moles acetyl-CoA to 2 moles ethanol, sustaining glycolytic flux¹⁵⁵. A prerequisite for PDHC-AdhE activation is the inactivation of the PflB reaction. Although PflB is oxygen-sensitive, its activity in lysates and aerobic environments has been reported^{74,156,157}. Active PflB converts 1 mole pyruvate and 1 mole CoA to 1 mole acetyl-CoA, forming 1 mole formate in the process. Because this reaction does not generate NADH, subsequent ethanol formation via NADH-dependent AdhE is imbalanced with glycolysis (**Figure 4.1A**)^{74,109,155}.

The post-lysis affinity-based enzyme depletion method was optimized to remove 6xHis-tagged PflB and LdhA proteins from derived lysates to the point of inactivating these reactions, ensuring flux redistribution towards the PDHC-AdhE route of ethanol formation. In prior work, lysates were incubated with beads washed from a volume of His-Pur Cobalt resin suspension that was 1/5th the lysate volume⁷⁴. This method however still resulted in some formate production and significant lactate accumulation after 24 hrs. Utilizing more Co^{2+} beads to bind 6xHis-tagged proteins in the lysate should help eliminate flux-competing reactions. To this end, Co^{2+} beads washed from varying volumes of resin suspension were mixed with a fixed volume of lysate derived from the genome engineered strain, and mixtures were incubated for 1 hour at 4°C under shaking conditions. The resulting supernatants (i.e., depleted lysates) were normalized to the same protein concentration and used to prepare CFME reactions that were fed with a glucose substrate. All metabolite concentrations quantified in incubated reactions were normalized against time-zero controls, which were prepared as previously described¹⁵⁴. Further, the variability of the depletion approach was assessed by performing this method in triplicate from the same batch of lysate, then using the resulting “depletion replicates” to prepare replicate cell-free reactions (**Figure 4.2**). Despite batch variability being an issue in other lysate-based applications (e.g., protein expression)¹⁵⁸, metabolic performances of extracts prepared from biological replicates (i.e., different cell cultures) do not significantly vary^{72,74}.

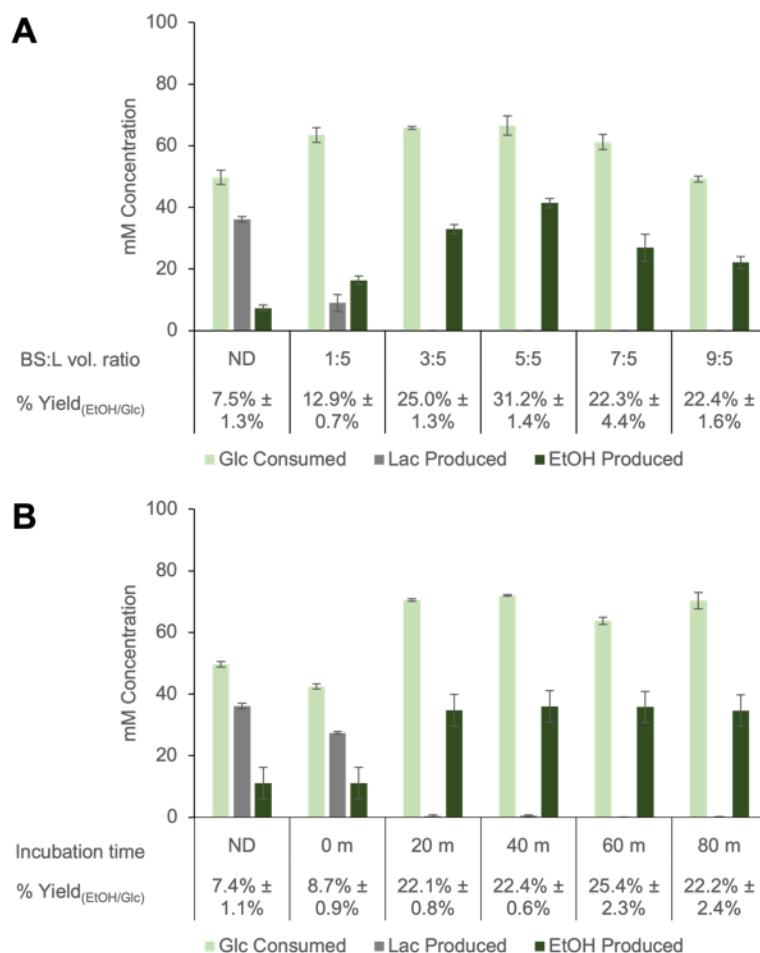


Figure 4.2. Optimization of the post-lysis lysate depletion method to selectively remove 6xHis-tagged PflB and LdhA and redirect lysate metabolic flux to ethanol. Normalized metabolite concentrations and calculated % ethanol yields in reactions made with **A)** lysates that were exposed to varying amounts of Co^{2+} beads for 1 hour, and **B)** lysates depleted with Co^{2+} beads washed from an equivalent suspension volume for different incubation times. All reactions were set up using 1 mM of NAD^+ , CoA, and ATP and 100 mM glucose. Reactions were incubated at 37°C for 20 hrs. Replicates of the depletion method (i.e., “depletion replicates”) were used to prepare triplicate reactions for all conditions and error bars represent standard errors. “BS:L vol. ratio” = bead suspension to lysate volume ratio. “ND” = non-depleted control.

Contrary to the previous report⁷⁴, no formate was detected in any of the modified extracts nor in the unmodified extract. PflB activity in lysates has been described in the past, so it is likely that formate was only accumulated here at undetectable levels or was actively consumed by formate dehydrogenases^{74,121,154}. Since it has been previously shown that LdhA and PflB can be simultaneously purified from the lysate by this procedure, decreasing lactate build-up was used as a proxy for evaluating the purification of both proteins⁷⁴. Reactions prepared with lysates exposed to an equivalent volume of bead suspension resulted in no lactate accumulation and the highest % ethanol yield from glucose ($31.2\% \pm 1.4\%$). This is a significant 2-fold improvement ($pval < 0.01$) compared to the % yield in the 1:5 volume ratio condition ($12.9\% \pm 0.7\%$), which was the only ratio tested in the previous study⁷⁴ (**Figure 4.2A**). This condition optimally enabled the removal of byproduct forming activity while increasing flux through the PDHC route of acetyl-CoA formation. Notably, exposure to even more beads (7:5 and 9:5 volume ratios) lowered glucose consumption, ethanol production, and % ethanol yield relative to the 5:5 condition (**Figure 4.2A**). This is possibly because of increased binding of background proteins in the lysate to the Co^{2+} beads⁷⁴. Nonetheless, all lysates processed using the 6xHis-tag purification strategy consumed more glucose and produced more ethanol when compared against a non-depleted (ND) lysate (**Figure 4.2A**). The effect of removing lactate-forming activity on the target yield increase is obvious in these data. To confirm the contribution of PflB depletion, the 5:5 volume ratio condition was also applied to lysates derived from a strain that endogenously expresses 6xHis-tagged LdhA only. Evidently, depleting both PflB and LdhA has a more positive impact on ethanol yield compared to LdhA depletion alone (**Appendix Figure 4.9**).

To determine whether increasing the exposure period of lysates to beads would improve selective enzyme depletion, lysates were mixed for varying durations with beads washed from an equivalent His-Pur Co^{2+} resin suspension and then used to prepare CFME reactions (**Figure 4.2B**). No significant improvement in % yield beyond 20 min of incubation was observed (**Figure**

4.2B). The results also suggest that potential shear forces from shaking the beads vigorously in lysates have no effect on glucose metabolism or % ethanol yield, otherwise lysate performance could be expected to decrease with time (**Figure 4.2B**). Moving forward, a standard 30 min incubation period of lysates with beads from an equivalent volume of Co^{2+} resin suspension was used to prepare modified lysates. Extracts depleted under these conditions are henceforth referred to as *blocked* lysates.

The reaction blocking strategy described here has several advantages over other tactics to rewire lysate metabolic flux. Gene knockouts can have unpredictable effects on harvested lysate metabolism, presumably because the source strain reconstructs its proteome to accommodate the deletion(s)^{47,53,159,160}. These mutant lines would result in lysate proteomes that differ on a global scale when compared to the wild-type¹⁶⁰. Intuitively, tractable changes to the “static” lysate proteome can be made by selectively removing enzymes post-lysis, enabling engineering designs with predictable outcomes and significantly simplifying the integration of lysate metabolic models. Displaying 6xHis-tags on endogenously expressed enzymes does not impact cell growth, whereas the number of gene knockouts that can be introduced without strain optimization is limited⁴⁷. Additionally, simply reducing the concentrations of flux-competing enzymes by down-regulating gene expression does not eliminate unwanted reactions in lysates^{62,74}. The strategy presented here already provides > 3-fold improvements in both titer and % yield compared against the unmodified extract even without optimization of the reaction environment or heterologous enzyme expression, both of which were necessary for achieving a similar improvement in BDO titer via a knockdown strategy⁶².

4.4.2. Optimized reaction and cultivation conditions push and pull flux through the insulated pathway and improve ethanol yield in blocked lysates

Removing enzymes alone can be insufficient for achieving high conversion yields. Given the complexity of lysate metabolism, glucose-derived carbon can be trapped in metabolic intermediates or incorporated into less obvious end-products of central metabolism¹⁵⁴. Push and pull metabolic engineering strategies are typically implemented to improve pathway yields by overcoming bottleneck reactions while increasing carbon flow towards a target metabolite¹⁴⁸. Previous CFME efforts have shown that optimal reaction and source strain cultivation conditions can enrich the concentrations or activities of desirable enzymes in lysate-based systems, increasing flux through a pathway of interest^{17,19,53,62,73,78}. Some of these tactics were thus applied sequentially to increase flux through the insulated ethanol biosynthesis pathway in the blocked lysate, using % yield of ethanol from glucose as a parameter for optimization.

First, different combinations of starting cofactor concentrations were sampled in reactions prepared with glucose fed, blocked extracts. The cofactors NAD⁺, ATP, and CoA are all responsible for driving past flux restricting steps in the glycolysis module of the target pathway (**Figure 4.1A**), hence the initial supplementation of 1 mM of each of these reagents to the reactions. Interestingly, ~ 40% ($39.1\% \pm 3.2\%$) of the maximum theoretical ethanol yield can be achieved when the reactions are started with 2 mM NAD⁺ (**Figure 4.3A**); supplementation with ATP or CoA are unnecessary for improving yield. These conditions result in a % yield that is significantly higher ($p\text{-val} = 0.03$) than that of the initial condition ($27.2\% \pm 3.5\%$) (**Figure 4.3B**). NAD⁺ concentrations beyond 2 mM had no positive effect on ethanol yield (**Appendix Figure 4.10**). Approximately 40% yield could also be realized if 1 mM ATP and 1 mM NAD⁺ were added to reactions without CoA, suggesting a complex interplay between ATP and NAD⁺ availability (**Figure 4.3B**). Regardless of starting ATP or NAD⁺ concentrations, decreasing CoA concentrations consistently improved % ethanol yields (**Figures 4.3A and B**). This can be

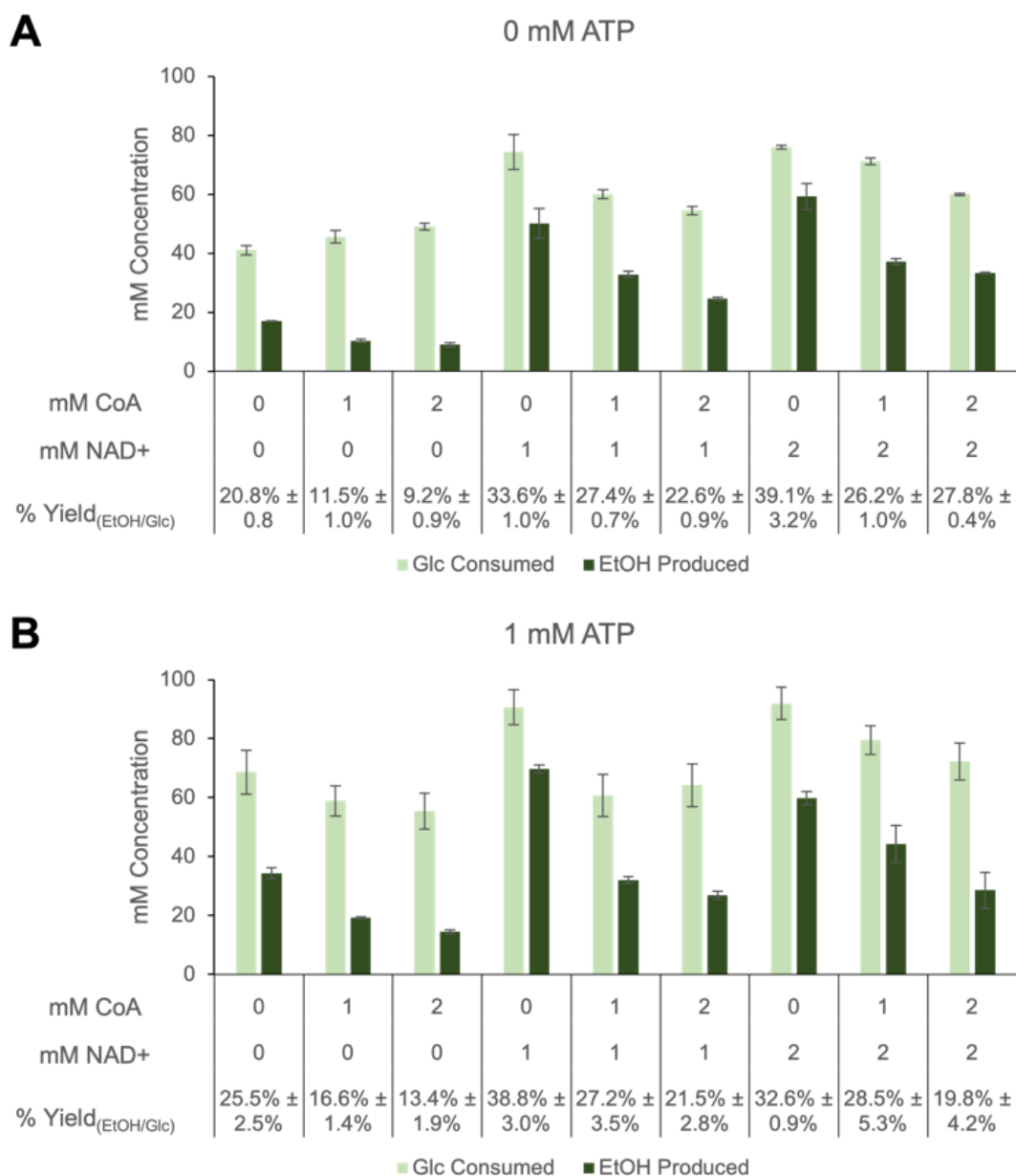


Figure 4.3. Optimization of cofactor concentrations in reactions prepared with blocked lysate to drive carbon flux towards ethanol. Normalized metabolite concentrations and calculated % ethanol yields in reactions of blocked lysates prepared with 100 mM Glucose, different combinations of 0, 1, or 2 mM of NAD⁺ and CoA, and **A)** 0 mM or **B)** 1 mM ATP. Reactions were incubated at 37°C for 20 hrs. Reactions for all conditions were prepared using depletion replicates and error bars represent standard errors (n=3).

attributed to the regulatory role of CoA-derived moieties which can bottleneck flux through glycolysis to ethanol. When accumulated, succinyl-CoA and acetyl-CoA inhibit the pyruvate forming enzyme pyruvate kinase and PDHC, respectively^{109,161,162}. These intermediates can become excessive when there is high CoA availability at the start of the reaction and result in lower glucose consumption and ethanol production, thereby lowering % ethanol yields (**Figures 4.3A and B**). Another possibility is that free CoA inhibits the forward reaction catalyzed by the acetaldehyde dehydrogenase component of AdhE, stalling ethanol synthesis¹⁶³.

Next, different reaction vessels were tested to determine whether smaller sized reactors would benefit the system's conversion efficiency. Decreasing the size of the vessel reduces the air/liquid surface area to reaction volume ratio as well as the reaction headspace, both of which reportedly limit the oxygenation of cell-free reactions^{164,165}. In *E. coli* cells, the electron transport chain and component oxygen-dependent enzymes that regenerate NAD⁺ from NADH are inactivated when oxygen is limited^{109,166–168}. Increased [NADH]/[NAD⁺] in the cytosol is consequently regulated by the increased activity of endogenous AdhE enzymes^{166,168}, a phenomenon that also likely occurs in oxygen-limited *E. coli* lysates¹⁶⁹. Additionally, a smaller air/liquid surface area and reaction headspace could better prevent ethanol evaporation from the reaction volume. Hence, 30 μ L glucose-fed reactions prepared with blocked lysates and optimized cofactor concentrations were incubated in reaction vessels of varying sizes: 1.5 mL microcentrifuge tubes, 0.2 mL PCR tubes, and 50 μ L non-binding surface wells in a 384-well plate sealed with film. The air/liquid surface area to volume ratios of the 30 μ L reactions in these vessels were approximately 0.65 mm, 0.42 mm, and 0.32 mm, respectively. Expectedly, 30 μ L reactions in 50 μ L wells had a significantly higher mean % yield ($59.1\% \pm 1.4\%$) compared to the average $\sim 40\%$ yield of reactions in 1.5 mL microcentrifuge tubes ($pval < 0.01$) (**Figure 4.4A**).

Next, different reaction vessels were tested to determine whether smaller sized reactors would benefit the system's conversion efficiency. Decreasing the size of the vessel reduces the air/liquid surface area to reaction volume ratio as well as the reaction headspace, both of which

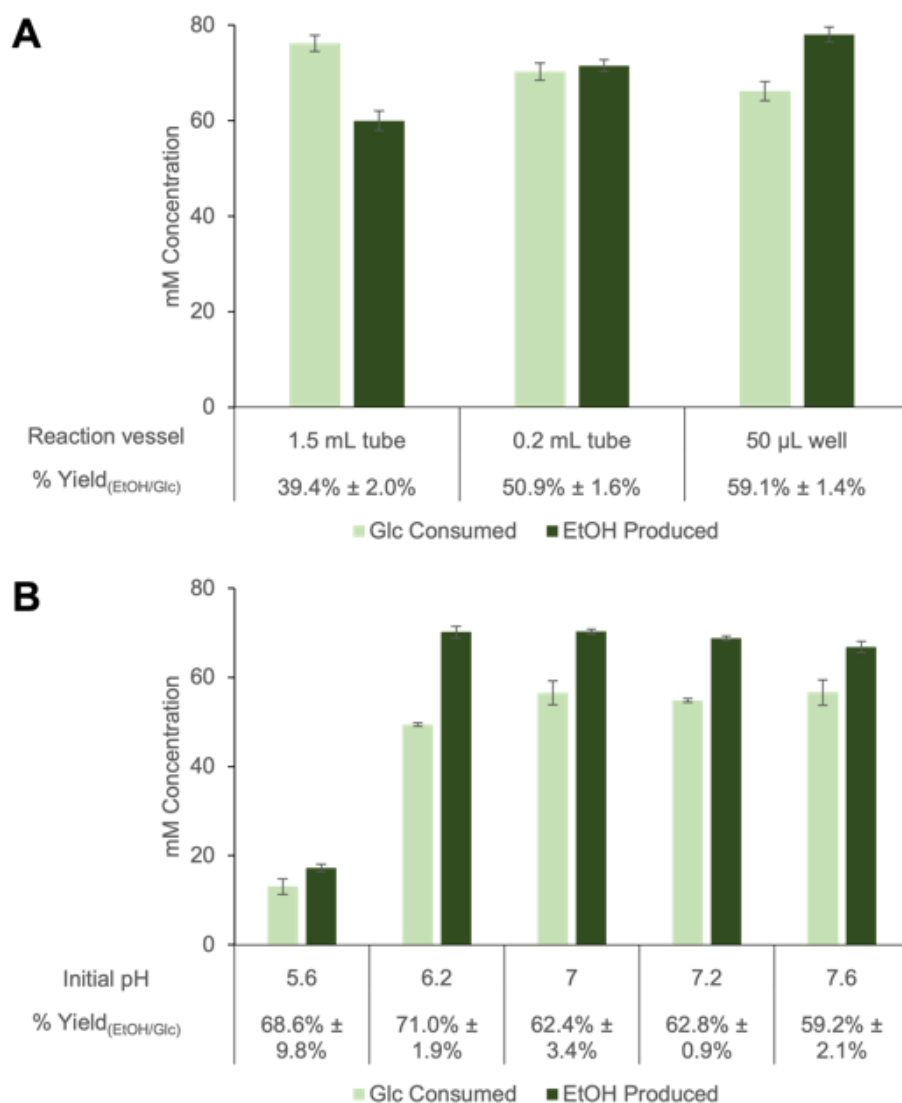


Figure 4.4. Optimization of reaction vessel and initial reaction pH to drive carbon flux towards ethanol in blocked lysates. Normalized metabolite concentrations and calculated % ethanol yields in reactions incubated in **A)** different enclosed reaction vessels with varying airspace volumes, and **B)** sealed 50 µL wells with varying initial pH values. Reactions were supplemented with 2 mM NAD⁺ and 80 mM glucose then incubated at 37°C for 20 hrs. Reactions for all conditions were prepared using depletion replicates and error bars represent standard errors (n=3).

reportedly limit the oxygenation of cell-free reactions^{164,165}. In *E. coli* cells, the electron transport chain and component oxygen-dependent enzymes that regenerate NAD⁺ from NADH are inactivated when oxygen is limited^{109,166–168}. Increased [NADH]/[NAD⁺] in the cytosol is consequently regulated by the increased activity of endogenous AdhE enzymes^{166,168}, a phenomenon that also likely occurs in oxygen-limited *E. coli* lysates¹⁶⁹. Additionally, a smaller air/liquid surface area and reaction headspace could better prevent ethanol evaporation from the reaction volume. Hence, 30 μ L glucose-fed reactions prepared with blocked lysates and optimized cofactor concentrations were incubated in reaction vessels of varying sizes: 1.5 mL microcentrifuge tubes, 0.2 mL PCR tubes, and 50 μ L non-binding surface wells in a 384-well plate sealed with film. The air/liquid surface area to volume ratios of the 30 μ L reactions in these vessels were approximately 0.65 mm, 0.42 mm, and 0.32 mm, respectively. Expectedly, 30 μ L reactions in 50 μ L wells had a significantly higher mean % yield ($59.1\% \pm 1.4\%$) compared to the average $\sim 40\%$ yield of reactions in 1.5 mL microcentrifuge tubes ($pval < 0.01$) (**Figure 4.4A**).

Cell-free systems provide the advantage of directly altering the reaction environment's pH, which can potentially improve activity of specific enzymes and therefore pathway performance. Sampling different initial reaction pH values has therefore become a common CFME strategy for improving product titers^{51,62,78,170}. This strategy was employed here specifically to improve ethanol yield in the blocked extract. *E. coli* cells regulate cytosolic pH within a range of 7.2 to 7.8, hence the use of pH 7.2 as the initial condition¹⁷¹. However, reducing the pH from 7.2 to 6.2 resulted in a significant % yield increase ($pval < 0.01$) to $71.0\% \pm 1.9\%$ (**Figure 4.4B**). This is attributed to a modest increase in ethanol concentration ($pval > 0.05$) and a significant decrease in glucose consumption ($pval < 0.01$) (**Figure 4.4B**). In *E. coli* cells, a reduction from cytosolic pH 7.0 to 6.0 leads to a 20% decrease in glycolytic flux¹⁷². Although the pH optima of enzymes *in vitro* are generally lower than that observed *in vivo*, it is conceivable that cell-free glycolytic enzymes are already slightly impaired at pH 6.2, leading to the reduced usage of glucose at this pH¹⁷³.

Intriguingly, however, the decrease in glucose consumption had no significant effect on ethanol concentration compared to the initial condition (**Figure 4.4B**). This suggests that carbon flux from consumed glucose is better directed towards ethanol in lysates with a slightly more acidic reaction environment. Further reducing the reaction pH results in a similar ethanol yield ($68.6\% \pm 9.8\%$) but compromises ethanol concentration (**Figure 4.4B**). For these reasons, pH 6.2 was considered the optimum for succeeding experiments.

To further improve ethanol yield, different source strain growth conditions were evaluated for enriching throughput of the target pathway. As reported, the lysate's endogenous metabolic proteome can be curated by growing cells under different nutritional conditions or by harvesting cells at different growth stages, improving CFME outcomes^{72,73}. Ethanologenic metabolism in *E. coli* cells is differentially expressed in cultures with varying concentrations of carbon substrate and growth phase^{167,174–176}. Glucose rich media stimulate glycolysis, but substrate-derived carbon is primarily funneled to lactate generation instead of ethanol synthesis under high glucose concentrations¹⁷⁴. *E. coli* ethanol production is known to occur during fast growth on glucose even under aerobic conditions, but up-regulated AdhE expression has also been documented in *E. coli* cultures at the onset of stationary phase^{175,176}. The glucose concentration for growth and harvesting time could thus theoretically be optimized to obtain lysates with enriched ethanologenic activity. To this end, the strain with genomically integrated 6xHis-tags was grown in 2xYPT media containing different concentrations of glucose. Cells were then pelleted for lysis at mid-log growth (OD ~ 5.0) and the onset of the stationary phase (OD ~7.0). Blocked extracts were derived from these cells and were fed glucose in reactions using the optimized reaction conditions. A lysate with high ethanol yield from glucose ($90.8 \pm 5.8\%$) could be extracted from cells grown exponentially on 0.9% glucose (**Figure 4.5**). Compared to lysates that were initially harvested from mid-log cells grown on 1.8% glucose, the optimized lysate produced ethanol at significantly higher concentrations ($pval < 0.01$), leading to the significantly higher % ethanol yield from

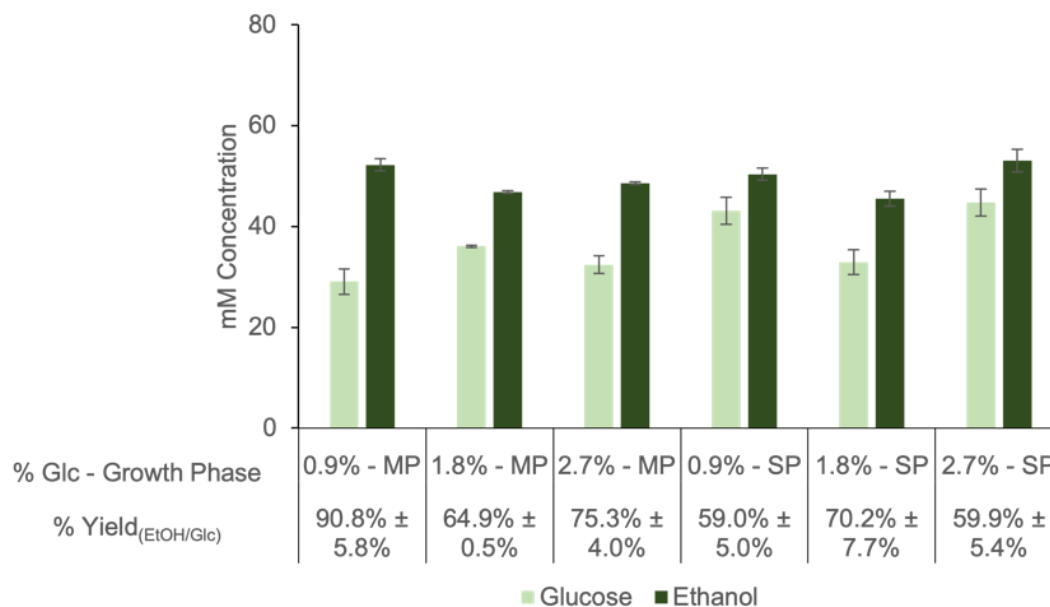


Figure 4.5. Optimization of source strain cultivation conditions to drive carbon flux towards ethanol in resulting lysate proteomes. Normalized metabolite concentrations and calculated % ethanol yields in blocked lysates derived from strains grown under different conditions. Reactions were prepared with optimized reaction conditions and 80 mM glucose then incubated at 37°C for 20 hrs. Reactions for all conditions were prepared using depletion replicates and error bars represent standard errors (n=3).

glucose ($pval < 0.01$) (**Figure 4.5**). This > 90% theoretical yield is considered industrially relevant for ethanol biosynthesis^{20,146}.

In efforts to maximize redirection of glucose-derived carbon flux towards ethanol synthesis in crude lysates, reaction conditions defined as optimal should yield high glucose-to-ethanol transformation while minimizing acetate formation or consumption. In addition to ethanol production, changes in the system's acetate concentrations were analyzed throughout optimization experiments. Both the production and consumption of acetate in *E. coli* lysate-based cell-free systems have been reported^{46,53,74,154}. Enzymes involved in the acetyl-CoA—acetate pathway can drive reactions in either direction depending on the lysate's adenylate charge¹⁵⁴. That is, the pathway, which involves ATP/ADP exchange, can either form or deplete acetate depending on the availability of these adenosine cofactors (**Figure 4.1A**). When this pathway is driven forward, acetate is formed alongside ethanol since the two fermentative end-products share the same acetyl-CoA precursor. Consequently, acetate production is likely to compete with ethanol synthesis for glucose-derived carbon flux¹⁷⁷. When driven backward towards acetyl-CoA synthesis, acetate can be a secondary carbon source for ethanol formation since it is initially present in the reactions at a final concentration of approx. 80 mM⁵³. This is because the S30 buffer used to prepare lysates consists of acetate salts. If acetate is indeed consumed in the reactions, this would obscure % ethanol yield from glucose calculations. When optimizing for starting cofactor concentrations, acetate was observed in the blocked extract. Like ethanol, acetate increased with decreasing CoA, reinforcing the assumption that glycolysis is bottlenecked by high CoA levels (**Appendix Figure 4.11A**). Incubating reactions in more oxygen-limited environments however increasingly drove the acetyl-CoA—acetate pathway module towards acetyl-CoA synthesis, as evidenced by acetate consumption in 50 μ L vessels (**Appendix Figure 4.11B**). This could be explained by the need for an additional ATP consuming sink that is fulfilled by the acetate kinase (AckA) reverse reaction when O₂-dependent ATP hydrolases are inactivated in the system^{154,164}. Despite acetate consumption in the sealed 50 μ L wells, data from

the initial reaction pH optimization experiment imply that carbon derived from this secondary source is likely not incorporated into ethanol since ethanol concentrations do not significantly change in conditions where acetate is produced (pH 6.2) or consumed (pH > 7.0) (**Appendix Figures 4.11C & B**). These data also reinforce the use of pH 6.2 for ethanol production in lysates since this condition has the least significant acetate production or consumption. Under these reaction conditions, acetate formation or consumption remained insignificant in extracts regardless of source strain cultivation conditions (**Appendix Figure 4.11D**).

To further ensure that ethanol in this optimized system was formed from glucose, reactions (n=2) were incubated for 20 h without glucose and compared to respective time zero controls. No glucose or ethanol was detected in any of the time zero replicates and only 1.9 ± 0.08 mM ethanol was formed after 20 h. This strongly suggests that ethanol is derived from glucose, and not from carbon sources initially present in the reactions at high (> 80 mM) concentrations (i.e., acetate and glutamate)⁴⁹.

4.4.3. Maximum glucose-to-ethanol conversion is consistent in the optimized system

The cell-free, block-push-pull approach enabled a 10-fold improvement in the optimized system's ethanol yield from glucose compared to the unmodified extract (**Figure 4.6**). The consistency of the optimized system to perform at > 90% yield was then assessed.

First, it was determined whether flux is still driven to ethanol without high concentrations of remaining substrate. In reaction optimization experiments, only a fraction of the glucose feed (~ 25 mM out of 80 mM) is consumed after 20 h (**Figure 4.5**). By Le Chatelier's principle, the high remaining concentration of glucose could be favoring the overall forward reaction to ethanol, preventing bottlenecks and build-up of intermediates. However, optimized reactions fed increasing concentrations of glucose convert a common concentration, 20-25 mM, of glucose to

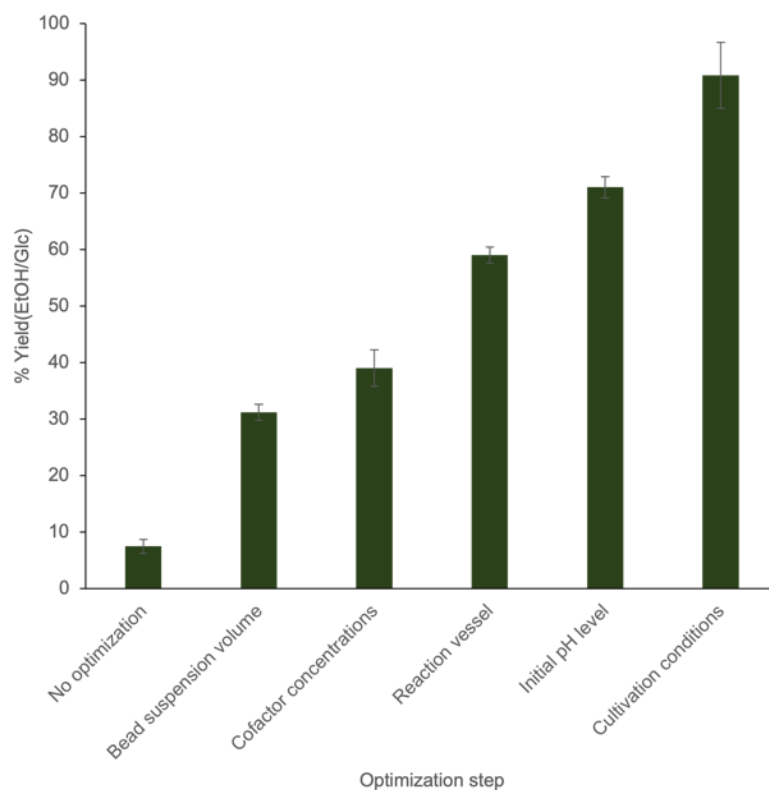
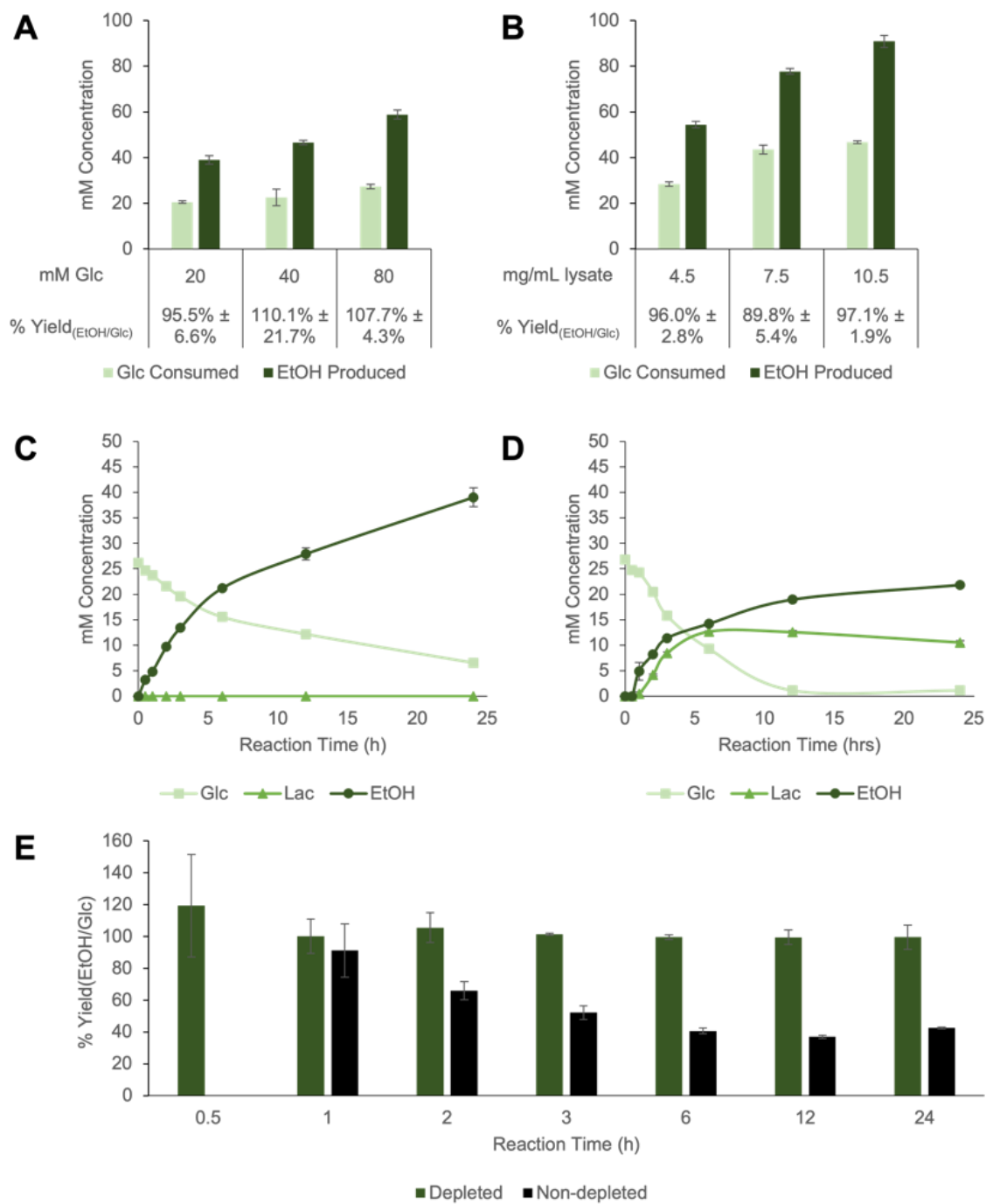


Figure 4.6. Summary of the additive improvements to % ethanol yield in the system due to conditions that optimally block, push, and pull flux. Reactions for all conditions were prepared using depletion replicates and error bars represent standard errors (n=3).

to ethanol at near 100% yields ($95.5\% \pm 6.6\%$ with 20 mM initial Glc; $110.1\% \pm 21.7\%$ with 40 mM initial Glc; $107.7\% \pm 4.3\%$ with 80 mM initial Glc) (**Figure 4.7A**). Reactions are thus driven forward towards ethanol synthesis, even without excess glucose in the system, suggesting that glucose consumption, and therefore ethanol concentrations, are limited by another component of the reaction system. The total concentration of relevant pathway enzymes was hypothesized to limit the reaction. To this end, reactions made with increasing concentrations of blocked lysate were fed 50 mM glucose and incubated overnight. Compared to the initial reaction condition containing 4.5 mg/mL lysate, reactions with more concentrated extracts could consume more glucose and produce more ethanol while maintaining near 100% yields ($96.0\% \pm 2.8\%$ with 4.5 mg/mL lysate; $89.8\% \pm 5.4\%$ with 7.5 mg/mL lysate; $97.1\% \pm 1.9\%$ with 10.5 mg/mL lysate). (**Figure 4.7B**). Thus, increasing the concentration of extract in the mixture can be a useful strategy for improving ethanol production and possibly even the system's productivity without impacting yield. However, it is also conceivable that the positive effect of concentrating extract on reaction production efficiency has an upper limit due to reaction viscosity^{178,179}. Alternatively, terminal enzymes may be overexpressed heterologously to help drive flux through the insulated pathway^{46,62}.

To confirm that these highly efficient conversions were consistent over time, a time course experiment was set up with 4.5 mg/mL lysate in reactions that were fed 25 mM glucose, the maximum amount of substrate this system would consume (**Figure 4.7A**). Glucose consumption inversely correlated with ethanol production, resulting in 100% ethanol yield in the blocked lysate at every timepoint (**Figure 4.7C & E**) These results strongly suggest that as glucose is consumed, its carbons are fully shuttled and incorporated into ethanol in the optimized system. Additionally, no significant production or consumption of acetate was observed through time (**Appendix Figure 4.12**). Furthermore, reactions prepared under the same reaction and cultivation conditions, but with non-depleted extract, accumulated half as much ethanol and only transformed glucose to

Figure 4.7. Evaluation of maximal glucose to ethanol conversion efficiency in the optimized system. Normalized metabolite concentrations and calculated % ethanol yields in the optimized reaction system prepared with **A)** 4.5 mg/mL lysate and different concentrations of glucose feed or **B)** 50 mM glucose and varying concentrations of lysate. Normalized metabolite concentrations in time course reactions set up with **C)** the system optimized via the block-push-pull approach and **D)** its non-depleted lysate counterpart (i.e., a system prepared with the optimized reaction and cultivation conditions but without removing competing reactions). These reactions were fed 25 mM glucose. **E)** Percent ethanol yields in the optimized system and its non-depleted lysate counterpart through time. All reactions were incubated at 37°C. Except for the time course experiment, reactions were run for 20 hrs. Reactions for all conditions were prepared using depletion replicates and error bars represent standard errors (n=3).



ethanol at less than half of the theoretical yield in 24 h. (**Figure 4.7D & E**) These data reinforce that removing unwanted reactions was indeed necessary for effectively rewiring flux. However, this still had to be complemented by strategies to drive flux through bottlenecks and towards ethanol to realize the maximum yield. This is implied by the ~ 50% yield obtained in the non-depleted system under optimized cultivation and reaction conditions (**Figure 4.7E**), which is a ~ 5-fold improvement from the < 10% yield in the non-optimized system (**Figure 4.6**).

It should be noted that the non-depleted lysate consumed glucose more rapidly than the blocked extract (**Figures 4.7C & D**). At early timepoints (1, 2, 3 h), this seems to be due to the lactate-forming ability that helps AdhE turn over excess NADH to drive rapid glycolytic flux (**Figure 4.7D**). Indeed, > 90% of glucose carbon was incorporated into lactate and ethanol at these earlier timepoints. Interestingly, when remaining glucose was completely consumed in the non-depleted extract between 6 and 12 h, less glucose carbon was incorporated into ethanol, lactate, or other fermentative products that are detectable by the HPLC method used here (**Figure 4.7D**). Whether this is because flux is being directed to non-fermentative end-products or because bottlenecks are causing the build-up of intermediates requires further investigation. Regardless, this finding highly contrasts the continuous production of ethanol from glucose with > 90% theoretical yield beyond 12 h in blocked extracts, further supporting that the workflow efficiently directs carbon flux from glucose to ethanol (**Figure 4.7E**).

4.5. Conclusions

Here, a lysate-based cell-free system was optimized to synthesize ethanol from consumed glucose with an industrially relevant % yield (> 90%)¹⁴⁶. To our knowledge, this is the first report of metabolic flux rewiring in lysates that maximizes the yield of a targeted compound. To accomplish this, a block-push-pull approach was designed and implemented. Specifically, an efficient strategy for depleting flux-competing reactions from the lysate itself was first employed.

CFME methods that could drive flux through the insulated target pathway were then applied to improve the system's conversion of substrate to product. These strategies take advantage of the manipulable cell-free environment. That is, the lysate proteome could be curated without introducing metabolic burden from protein overexpression or deleterious mutations, and pathway activity could be rapidly modulated by directly manipulating reaction conditions. The engineering workflow here is therefore unconstrained by cell viability and enables shorter build steps in DBTL cycles. As a result, the yield of the target endogenous metabolite could be quickly maximized in a lysate which initially synthesized ethanol from glucose at only 10% of the theoretical yield. A source strain that already had the native or introduced capacity to synthesize the targeted product with high performance metrics was therefore not needed^{62,71,180}. This implies that time-consuming strain optimizations or the availability/discovery of already optimized strains to rewire endogenous metabolic flux are not required for engineering lysates that result in high conversion yields. This can be advantageous for reducing the number of heterologous genes introduced into the source strain, reducing manipulations to only those that are necessary for completing cell-free pathway designs. Elements of the current workflow can be feasibly applied at the bench-scale to study cellular metabolism, test unconventional phenotypes, or prototype potential high-yield pathway designs for *in vivo* production. It is important to note that the depletion strategy is reliable for such applications as it does not cause significant variability. However, the current method could benefit from testing other tag-resin combinations with higher specificity, and an assessment of the number of enzymes that can be feasibly depleted by these approaches.

Efficient bioconversion is essential to realizing lucrative lysate-based biomanufacturing. According to estimations provided by Rasor et al., the synthesis of certain high-value molecules (e.g., limonene, 3-hydroxybutyrate) can be profitable in cell extracts provided that available reaction cost reduction strategies are integrated, and maximum conversion yields can be engineered in lysate pathways¹¹. The latter is demonstrated by the optimizations described here, which were designed specifically to propel flux through glycolysis for ethanol production.

Appropriately designed cell-free block-push-pull approaches should similarly be able to direct carbon flow to other target metabolites. CFME methods that enrich and activate enzymes of the target pathway, integrated into the described approach as push and pull strategies, have already been applied to increase titers and productivities of lysates that synthesize a broad range of compounds^{28,47,51,62,71,73,145}. If complemented with an efficient strategy for removing unwanted reactions, it is conceivable that high yields of these metabolites can also be achieved. However, refinements should be made to the described depletion approach to enable sustainable production in extracts due to the high cost of the metal chelate affinity-based strategy (**Appendix Note**). Other protein tag technologies could be considered. These tags must be small enough so that 1) their sequences can be easily inserted into genomes and 2) they do not affect the endogenous enzyme's physicochemical properties. Cellulose binding domains with sequences as short as 27 amino acids long¹⁸¹ can be utilized as protein tags and tested with inexpensive cellulose beads^{182,183}. We estimate that lysate depletion with cellulose beads can enable profitable lysate-based biomanufacturing in the future (**Appendix Note**). The *mf*-ssrA tag is also 27 amino acids in length and is sensitive to proteolysis by an exogenous *mf*-Lon protease¹³². This system was previously recommended for removing flux-competing reactions post-lysis⁷⁴ but the cost of large-scale protease purification should be analyzed. An alternative strategy is the use of a native protease in *E. coli* that only degrades tagged enzymes upon cell lysis¹⁸⁴. An OmpT protease cleavage site (a dibasic amino acid sequence) was engineered into a strain's endogenous cytosolic protein and cleaved by the native OmpT protease, which is localized in the outer membrane, only after membrane disruption. This method had no impact on cell growth and could thus be adapted as a block strategy for pathway yield optimization. However, this demonstration was focused on degrading a single protein for enhanced non-canonical amino acid incorporation in cell-free protein synthesis reactions¹⁸⁴. Using this approach to selectively lyse multiple enzymes for metabolic flux rewiring may introduce new challenges, like OmpT's potential proteolysis of metabolic enzymes with dibasic cleavage sites as OmpT reportedly has more degradation targets

in extracts^{38,185,186}. Non-native proteases like *mf*-Lon may be assessed for this type of approach, but tuning the heterologous expression of the outer membrane-localized protease would likely require strain optimization^{25,133}.

This work described here illustrates the feasibility of engineering high-yield bioconversion in complex lysate environments by optimizing and combining CFME approaches. Looking forward, we anticipate that the design and execution of pathway-specific, cell-free block-push-pull workflows that can achieve industrially relevant yield metrics for high value molecules in extracts. Variations of the described workflow can be leveraged to study cellular metabolism or to prototype pathway designs before construction *in vivo*. Innovative, low-cost strategies for selective depletion of enzymes will advance commercially viable lysate-based biomanufacturing systems, providing an alternative engineering platform for rapid and sustainable bioconversion.

Appendix

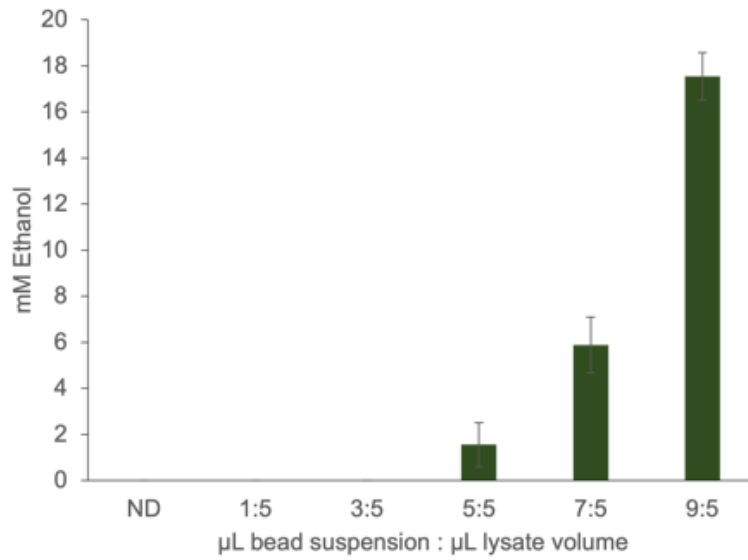


Figure 4.8. Residual ethanol from resin suspension storage buffer carried over to time zero lysate reaction set-ups.

Table 4.1. Number of cycles needed to wash ethanol off different resin suspension volumes with S30 buffer (for 200 μL lysate).

Volume resin suspension (μL)	Number of wash cycles
40	2
120	2
200	3
280	5
360	7

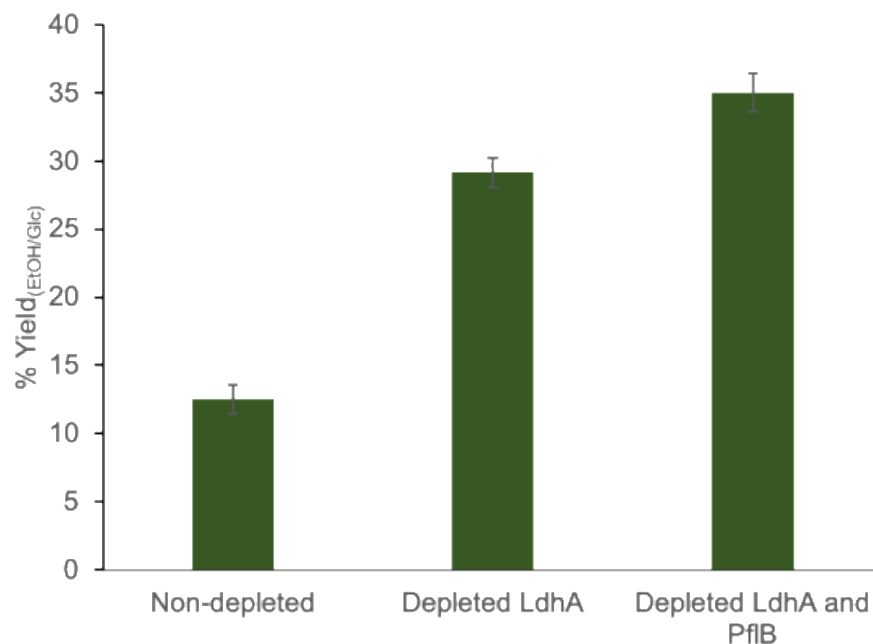


Figure 4.9. Contribution of PflB depletion to the improvement of ethanol yield in LdhA depleted lysates. Lysates were derived from a strain that endogenously expresses both 6xHis-tagged LdhA and PflB or 6xHis-tagged LdhA alone. Lysates were depleted by mixing 200 μ L lysate with washed Co^{2+} beads from a 200 μ L resin suspension. Reactions were initiated with 100 mM glucose and incubated for ~ 20 hrs at 37°C along with non-depleted control reactions. Depletion reactions were used for all conditions and error bars represent standard error of the mean (n=3).

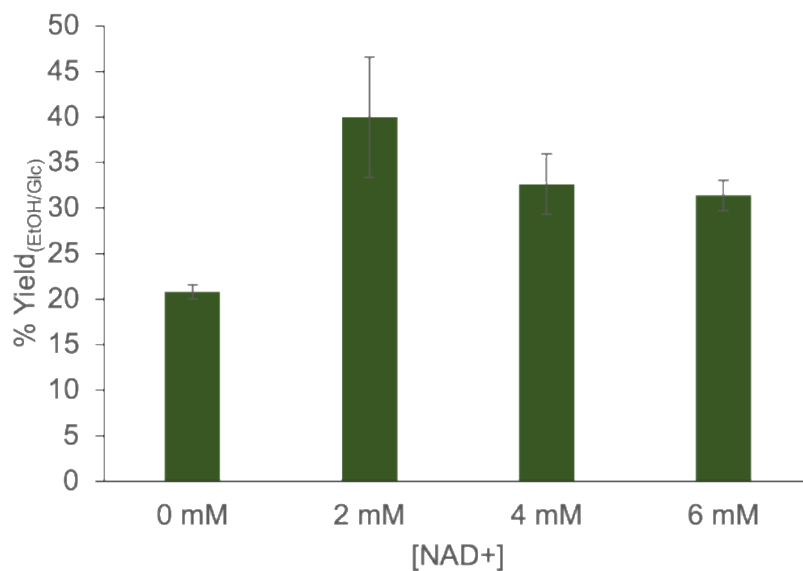


Figure 4.10. Additional NAD⁺ was supplemented to reactions prepared with blocked lysate, 100 mM Glucose, 0 mM CoA and ATP, but had no improvement or significant effect on % ethanol yield from glucose beyond 2 mM NAD⁺. Reactions were incubated at 37°C for 20 hrs. Depletion replicates were used for all conditions and error bars represent standard errors (n=3).

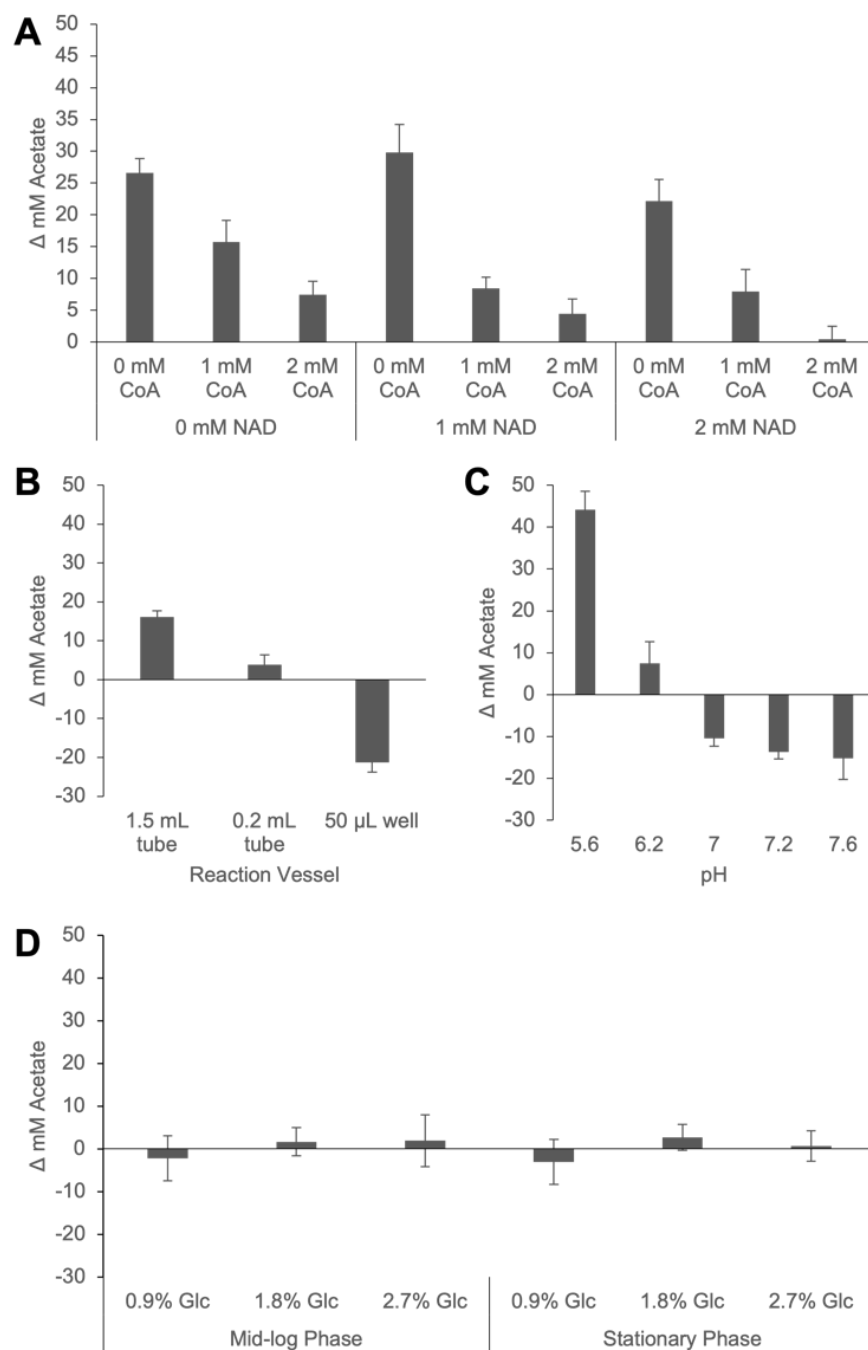


Figure 4.11. Acetate production or consumption in reactions. These were **A)** initiated with 0 mM ATP and various combinations of NAD⁺ and CoA, **B)** incubated in different enclosed reaction vessels, **C)** initiated with varying pH levels, **D)** prepared with lysates derived from strains grown in different cultivation conditions. Depletion replicates were used for all conditions and error bars represent standard errors (n=3).

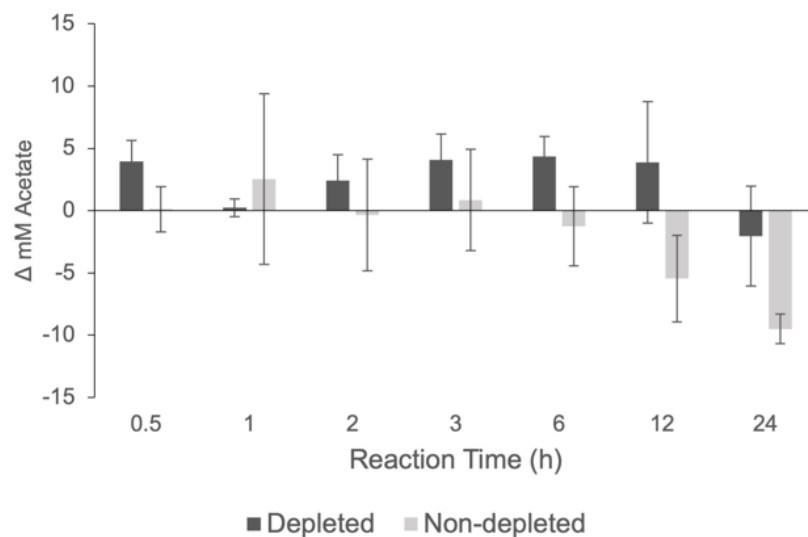


Figure 4.12. Acetate production or consumption in the optimized system and its non-depleted lysate counterpart. Values were calculated as the difference between acetate concentrations at a particular timepoint and at time zero reactions. Depletion replicates were used for all conditions and error bars represent standard errors (n=3).

Appendix Note:

The cost of the current selective enzyme depletion approach (i.e., block strategy) is evaluated using reaction preparation costs that were already calculated by Rasor et al¹¹. Briefly, they estimate the breakeven cost of a reaction that produces high-value targets (e.g., limonene and 3-hydroxybutyrate) to be \$275.00/L reaction. They define the breakeven reaction cost as “the cost of reaction required to start having a return on investment at a given max theoretical yield.” With innovative cost-reducing strategies (i.e., efficient cofactor utilization, single-carbon substrates, high-density cultivation) and neglecting labor/capital costs, the price of reaction preparation can be reduced to \$90.00/L reaction in the future.

In the current approach, equal volumes of lysate and Co²⁺ resin suspension (Thermo Scientific #89964) are needed to eliminate byproduct-forming activity. The resin is \$132.00/10 mL bottle. The resin itself therefore adds \$0.40 to the preparation of a 30 µL reaction. Assuming \$90.00/L reaction (\$0.0027/30 µL reaction), the total cost for preparing a system with the modified extract is \$0.4027/30 µL reaction or \$13,423.00/L reaction. Potential bulk pricing and other economies of scale may reduce these costs, but the present depletion strategy is unlikely to be viable for commercialization.

Alternative pre-purification strategies may enable sustainable yield engineering efforts. Islam et al. recently reported an approach involving the purification of proteins via cellulose binding domains (CBD) and microcrystalline cellulose (Sigma-Aldrich #310697)¹⁸². In their method, 700 mg of microcrystalline cellulose, which costs \$49/50 g, could be used to make 2000 µL bead volume. If each target endogenous enzyme is engineered to display a CBD and an equivalent volume of bead volume is needed to eliminate byproduct-forming activity, the material costs \$0.0010/30 µL. Assuming \$90/L reaction, the total cost for preparing a system with the modified extract is \$0.0040/30 µL reaction or \$133/L reaction. Even without considering economies of scale, this is well below the breakeven reaction cost for high-target chemicals. With

method optimization, it is possible that less microcrystalline cellulose is necessary to inactivate byproduct-forming activity, further reducing preparation costs.

CHAPTER 5

APPLYING A REPORTER SYSTEM TO INVESTIGATE AND OPTIMIZE THE LYSATE-BASED CELL-FREE EXPRESSION OF NONRIBOSOMAL PEPTIDE SYNTHETASES

A version of this chapter has been submitted for publication by Jaime Lorenzo N. Dinglasan, Tien M. Sword, William J. Barker, Mitchel J. Doktycz, and Constance B. Bailey.

Dinglasan, J.L.N.,* Sword, T.M.,* Barker, W.J., Doktycz, M.J., Bailey, C.B. (2023). Applying Reporters to Investigate and Optimize Lysate-based Cell-free Expression for Nonribosomal Peptide Production. *Accepted in ACS Synthetic Biology*. *Authors contributed equally

This chapter has been adapted from its published format to accommodate new Figure, Table, and Equation enumeration. The Supplementary Information associated with this work may be found in the Appendix section of this chapter. All references are located at the end of the document.

5.1. Abstract

Lysate-based cell-free expression (CFE) systems are accessible platforms for expressing proteins that are difficult to synthesize *in vivo*, such as nonribosomal peptide synthetases (NRPSs). NRPSs are large (> 100 kDa), modular enzyme complexes that synthesize bioactive peptide natural products. This synthetic process is analogous to transcription/translation (TX/TL) in lysates, resulting in potential resource competition between NRPS expression and NRPS activity in cell-free environments. Moreover, CFE conditions depend on the size and structure of the protein. Here, a reporter system for rapidly investigating and optimizing reaction environments for NRPS CFE are described. This strategy is demonstrated in *E. coli* lysate reactions using blue pigment synthetase A (BpsA), a model NRPS, carrying a C-terminal tetracysteine (TC) tag which forms a fluorescent complex with the biarsenical dye, FIAsh. A colorimetric assay was adapted for lysate reactions to detect the blue pigment product, indigoidine, of cell-free expressed BpsA-

TC, confirming that the tagged enzyme is catalytically active. An optimized protocol for endpoint TC/FIAsH complex measurements in reactions enables quick comparisons of full-length BpsA-TC expressed under different reaction conditions, defining unique requirements for NRPS expression that are related to the protein's catalytic activity and size. Importantly, these protein-dependent CFE conditions enable higher indigoidine titer and improve the expression of other monomodular NRPSs. Notably, these conditions differ from those used for the expression of super folder GFP (sfGFP), a common reporter for optimizing lysate-based CFE systems, indicating the necessity for tailored reporters to optimize expression for specific enzyme classes. The reporter system is anticipated to advance lysate-based CFE systems for complex enzyme synthesis, enabling natural product discovery.

5.2. Introduction

Nonribosomal peptides are a structurally diverse class of natural products with valuable biological activities that include antibiotics, immunosuppressants, insecticides, fungicides, and pigments among others¹⁸⁷. These short peptides are made independent of the ribosomal machinery, and are instead synthesized by large (> 100 kDa) enzyme complexes called nonribosomal peptide synthetases (NRPSs)¹⁸⁸. NRPSs are modular proteins and can exist as a single module or can be comprised of multiple modules organized into a complex assembly line. Each module has catalytic domains that carry out discrete chemical transformations to build a peptide product from amino acid precursors that are tethered by flexible linkers^{188–190}. Considerable efforts to reprogram NRPS architectures to generate novel peptides have been made, both in terms of altering selectivity of amino acid recognizing or loading domains through point mutations^{191–193} as well as domain swapping^{194–196}. However, many of these designs are hindered by slow experimental throughput including the ability to screen large enough numbers of constructs to develop adequate design rules^{67,193,196–203}. Additionally, there is a desire to study

uncharacterized NRPSs identified through genome mining which will enrich the portfolio of valuable nonribosomal peptide structures^{190,204}. To enable these endeavors, NRPS gene clusters have been expressed *in vivo* in microbial hosts¹⁸⁷. However, NRPS yields in cells are compromised by the metabolic burden from expressing such large and complex proteins, the cytotoxicity of their products, as well as the throughput limitations of genetics^{68,205}. Hence, there is increasing interest in developing platforms that allow high-throughput profiling of NRPS genes *in vitro* such as in cell-free expression (CFE) systems²³.

CFE platforms are considered highly suitable for NRPS production because they lack cellular survival objectives, bypassing challenges with *in vivo* expression^{28,68}. The open nature of CFE systems is additionally appealing for NRPS characterization and engineering because it can accelerate design-build-test-learn (DBTL) cycles. Direct manipulation of the expression environment affords high-throughput protein synthesis from exogenously added DNA as well as the introduction of non-natural components like non-canonical amino acids, all without requiring extensive cell engineering³⁵. To these ends, NRPSs have recently been expressed in both PURE²⁰⁷ and lysate-based^{54,208} cell-free systems. The nonribosomal peptide products made in these systems include the pigment indigoidine and rhabdopeptides in the PURE format, as well as valinomycin and a diketopiperazine precursor of gramicidin S in extracts²⁸. However, CFE is limited by components of the user-defined reaction environment, requiring the definition of effective reagent (e.g., energy substrates, salts, cofactors) concentrations that achieve sufficiently high protein expression yields^{35,64,158}.

Understanding the impact of a set of reaction conditions on CFE is a non-trivial task especially in lysate-based CFE systems which possess complex crosstalk between the lysate's active transcriptional/translational (TX/TL) and metabolic machineries²⁰⁹. For example, components meant to supply protein expression can also be metabolized or used to energize lysate metabolism in ways that unintentionally impact TX/TL processes²¹⁰. This can be especially complicated when synthesizing functional and structurally complex enzymes like NRPSs. The

catalytic domains of NRPSs could theoretically utilize reaction components that would otherwise sustain TX/TL, an additional resource competition problem in CFE systems that lacks prior investigation. For example, the NRPS adenylation (A) domain initiates NRP chain elongation by recognizing and activating an amino acid monomer as an adenylate using magnesium-ATP to generate an aminoacyl-AMP intermediate. This domain alone could thus contribute to unintentional lysate metabolism that competes for the TX/TL machinery's synthesis of the enzyme itself, including certain amino acids, the ATP producing energy substrate phosphoenolpyruvate (PEP), and magnesium ions. In addition to these active domains, the larger architectures of these proteins could be responsible for the production of NRPSs and NRPS-like proteins (i.e., polyketide synthases) complexes at low mg/L quantities and with high levels of premature truncation in lysate CFE systems^{54,211,212}. This is presumably because previous efforts have used reaction conditions that are not tailored for full-length translation of these large and complex enzymes^{54,211,212}, but instead were optimized to achieve g/L yields of small and robustly folding proteins such as the reporter proteins chloramphenicol acetyltransferase (CAT) and superfolder GFP (sfGFP)²⁰⁹.

Different reaction conditions can be sampled to probe limitations to NRPS CFE. Examining this parameter space however can be daunting due to barriers at the level of detection. That is, full-length NRPS CFE has only been validated through SDS-PAGE or mass spectrometric (MS) detection of the product, which by themselves can require time-consuming protocol optimization for different proteins or metabolites, limiting throughput^{54,211}. This feedback loop can be shortened with optically detectable peptide tags expressed at the C-termini of the expressed proteins²¹³. The use of a terminal tetracysteine (TC) peptide tag that forms a fluorescent complex with an exogenously added organoarsenic molecule has been utilized to monitor *in vitro* protein synthesis in the PURExpress cell-free system. However, its efficiency in the more complex cell extract environment has not been demonstrated²¹³. Adapting such reporters for lysate-based systems would have several advantages for NRPS CFE. On top of their relative accessibility and scalability compared to the PURExpress platform, lysate-based systems can be derived from strains that

already possess genetic components necessary for post-translationally modifying NRPSs into their functional *holo*-forms^{212,214}. Furthermore, alternate lysate systems have been developed for a range of other hosts including *Streptomyces sp.*^{215–218} and *Bacillus subtilis*²¹⁹, which are natural producers of industrially relevant NRPSs. Adapting reporters to the cell-free lysates would thus allow for applications in a larger range of transcriptional and translational machineries.

Here, the TC tag is used to investigate factors that limit full-length NRPS translation and, consequentially, metabolite production in an *E. coli* lysate-based cell-free system. Specifically, blue pigment synthetase A (BpsA, ~ 140 kDa), a model NRPS, with a terminal TC tag is expressed in *E. coli* lysate reactions^{220–222}. Based on measurements of BpsA's optically detectable pigment product, the cell-free expressed TC-tagged NRPS is functional in lysates. Furthermore, full-length expression of the tagged protein is detectable in backgrounds supplemented with FIAsH, a fluorogenic biarsenical dye which emits fluorescent signals when bound to the TC tag²²³. The TC/FIAsH reporter system can be used to sample the relative expression of full-length BpsA under different concentrations of PEP, magnesium, and amino acids, and to investigate how the NRPS's catalytic A domain competes against the lysate TX/TL machinery for these resources. Furthermore, conditions defined to improve BpsA synthesis based on TC-FIAsH measurements also increase BpsA activity, as suggested by improved pigment production. TC tag technology therefore can be leveraged as a strategy for quickly investigating and optimizing the lysate-based CFE of functional, full-length enzyme complexes like an NRPS²¹³.

5.3. Methods

5.3.1. Strains and plasmids

E. coli BL21 (DE3) was purchased from VWR (Radnor, Pennsylvania, USA). *E. coli* BAP1 was generously provided by Prof. Christopher Boddy (University of Ottawa, Ottawa, Canada) and has been described previously^{214,224}. *E. coli* BL21 Star (DE3) was purchased from New

England Biosciences (Ipswich, Massachusetts, USA). A pET28a(+) backbone containing the TC tag sequence, TGTTCGCCGGGTTGCTGT, and 6xHis-tag, CACCACCACCACCACCAC, was designed and ordered from TWIST biosciences (Table S1). The *bpsA* gene from *Streptomyces lavendulae* was codon optimized using Integrated DNA Technology's (IDT) codon optimization tool (<https://www.idtdna.com>) and was ordered as two gene blocks from IDT (part 1 and part 2) (Table S1) These were cloned into the described pET28a(+) vector using the GoldenGate assembly kit (New England Biolabs, part #E1601S) to generate the pET28a(+)-BpsA-TC plasmid. The codon-optimized *bpsA* sequence was mutated with the E315A substitution by mutagenic Gibson assembly (New England Biolabs, part #E2611S) to generate the pET28a(+)-BpsA-TC_E315A construct. The *pys* gene from *Pseudomonas fluorescens* and the *tycA* gene from *Bacillus brevis* were codon optimized and ordered as two gene blocks from IDT (part 1 and part 2) containing the TC tag sequence followed by a StrepII-tag, TGGAGCCACCCGCAGTTCGAAAAA (Table S1). The *pys* and *tycA* gene blocks were cloned into a pET28a(+) vector using the GoldenGate assembly kit (New England Biolabs, part #E1601S) to generate pET28a(+)-Pys-TC and pET28a(+)-TycA-TC plasmids. The plasmid pJL1-sfGFP was ordered from Addgene (part #69496). Primers used in this study were designed with the J5 software and are listed in Table S2.²²⁵

5.3.2. *In vivo* expression and purification of BpsA-TC

E. coli BAP1 was used as a host strain for *in vivo* BpsA-TC expression. Cultures were grown in LB (Miller) broth supplemented with kanamycin at 50 µg/mL. Overnight seed cultures (25 mL) were grown from a single colony at 37°C, shaking at 210 rpm. Flasks (VWR 29171-854) containing 1 L media were inoculated with 10 mL seed culture and incubated at 37°C, shaking at 210 rpm, until an OD₈₀₀ of 0.3-0.4 was reached. OD₈₀₀ rather than the more typical OD₆₀₀ was used to monitor growth to ensure that interference from the signal from indigoidine would not interfere with measurements of cellular turbidity. This strategy was adapted from Beer and

coworkers and was reproduced with the correlation between the OD₆₀₀ and OD₈₀₀.¹⁹⁸ The temperature was then lowered to 18 °C and cooled for 15 minutes prior to induction with 0.5 mM IPTG. The cultures were then incubated for ~ 20 hours at 18 °C prior to harvesting by centrifugation (4000 x g, 45 mins at 4 °C). Cell pellets were resuspended in wash buffer (5mM imidazole, 0.5 M NaCl, 10% v/v glycerol, 50 mM sodium phosphate, pH 7.8) and lysed by sonication. After sonication, the lysate was clarified via centrifugation (11,000 x g, 45 min, 4 °C) and purified via IMAC on an AktaPure system (Cytiva) using a 5 mL HisTrap column (Cytiva). Protein was eluted with elution buffer (400 mM imidazole, 0.5M NaCl, 10% v/v glycerol, and 50 mM sodium phosphate, pH 7.8) using a linear gradient from 20-60% elution buffer for 1 column volume (CV), 1 column volume (CV) of 60-100% elution buffer, and followed with a hold at 100% elution buffer for 3 CV. After purification, the collected fractions were dialyzed with 10K MWCO dialysis tubing (Thermofisher) for 2 passes for a minimum of six hours in storage buffer (50 mM phosphate pH 7.8, 100 mM NaCl, 10% (v/v) glycerol). The protein was then concentrated using a 100K MWCO concentrator (Thermofisher). Purity was assessed via SDS-PAGE using 4-20% pre-made gels (NuSep) (Fig S2).

5.3.3. *In vivo* indigioidine production and purification

A 25 mL seed culture of BAP1 harboring the pET28a(+)-BpsA-TC construct was grown overnight (37 °C, 2010 rev/min) in LB medium supplemented with 50 µg/mL kanamycin. After ~ 20 h, 10 mL of seed culture was used to inoculate 1 L media in a flask (VWR 29171-854). Cells were incubated at 37 °C shaking at 210 rev/min. At an OD₈₀₀ of 0.3-0.4, the culture was induced with 0.5 mM IPTG and grown at 16 °C for 20 h then lyophilized for 36 h. The dry cell mass was then washed with three rounds of DI water, methanol, ethanol, isopropanol, and hexanes to remove metabolites, salts, and proteins. Finally, the product was dried for one week under vacuum. Purity was verified via ¹H NMR (400 MHz, D6MSO) δ 11.30ppm (s, NH), 8.18ppm (s, CH), and 6.46ppm (s, NH₂) which is in agreement with previous literature report²²⁶.

5.3.4. Cell-free extract preparation

The same extract preparation procedure was used for all strains here. A seed culture was prepared with 30 mL 2xYPTG media (10 g/L yeast extract, 7 g/L potassium phosphate dibasic, 3 g/L potassium phosphate monobasic, 5 g/L NaCl, 16 g/L tryptone, and 18 g/L glucose) inoculated with a fresh colony and incubated overnight at 32°C, 220 rpm. 1 L of 2xYPTG media in a 2.5 L Tunair flask was then inoculated with the overnight culture and grown at 34°C, 220 rpm. Cell growth was monitored by NanoDrop (Thermo Scientific). At $OD_{600} \sim 0.9-1.0$, the culture was induced with 0.5 mM IPTG. Cells were harvested at $OD_{600} \sim 2.8-3.2$ by centrifugation (5000 x g, 15 min, 4°C), then washed three times using S30 buffer (10 mM Tris acetate, 14 mM magnesium acetate, 60 mM potassium acetate, and 10 mM DTT). All wash steps were performed at 4°C. Cell pellets were then weighed, flash frozen, and then stored at -80°C. For extract preparation, the cell pellets were then thawed on ice and resuspended in 0.8 mL of S30 buffer per g of cell pellet of the pellet by vortexing with short bursts (vortex 15 s, rest 30 s, repeat). 1.4 mL aliquots were sonicated on ice in 1.5 mL microcentrifuge tubes using a Branson 450 sonifier equipped with a 3.2-mm microtip (45 s on, 59 s off for three cycles, 50% amplitude set). 4 μ L of 1 M DTT was added into each tube immediately after sonication. All samples were centrifuged at 12000 x g for 10 minutes at 4°C. The supernatant was collected without disturbing the pellet and centrifuged again to remove remaining debris. The resulting supernatants were aliquoted into fresh centrifuge tubes, flash frozen, and stored at -80°C.

5.3.5. CFE reaction preparation

The initial cell-free reaction base mix comprised 1.2 mM ATP, 0.85 mM GTP, 0.85 mM UTP and 0.85 mM CTP; 34.5 μ g/mL folinic acid; 0.4 mM nicotinamide adenine dinucleotide (NAD), 0.27 mM coenzyme A (CoA), 4 mM Oxalic Acid, 1 mM, 1.5 mM spermidine, 57.33 mM HEPES buffer, 10 mM magnesium glutamate, 10 mM ammonium glutamate, 130 mM potassium glutamate, 2 mM each of the 20 amino acids, and 33 mM phosphoenolpyruvate (PEP). Reactions

were set up in 10 μ L volumes unless otherwise stated. Before condition optimization, reactions were started with 13 ng/ μ L DNA template and incubated at 30°C. The type of lysate used and changes to any of these conditions are described in the text. Unless otherwise stated, all CFE reactions were incubated for ~ 20 h. All reactions were incubated in a clear-bottom, black-wall 384-well plate with a non-binding surface (Corning #3544). Whenever possible, reactions were laid out in center wells. Surrounding wells were filled with 1x phosphate buffered saline (PBS) to control the humidity and prevent evaporation. Plates were covered with a hard covered lid and wrapped with parafilm at the sides before incubation.

5.3.6. Indigoidine quantitation in lysates with absorbance measurements

To generate standard curves from pigment absorbance measurements, increasing concentrations of the purified pigment dissolved in DMSO were spiked into BAP1 lysate mock reactions (i.e., reaction mix without DNA). Absorbance measurements were made in a 384-well plate, without a lid, loaded into a BioTek Cytation5 (Agilent) plate reader. The read protocol was set to shake the plate at high speed for 2 s then measure absorbance in selected wells at 590 nm and then at 800 nm. Absorbance values at 590 nm (A590) were divided by values at 800 nm (A800) following a previously described protocol in *E. coli* cells to deconvolute pigment absorbance from the lysate's optical density.¹⁹⁸ To solubilize purified indigoidine, 50 %, 60 %, 70 %, or 80 % DMSO were added to reaction mixes containing purified indigoidine following a protocol previously described in purified BpsA samples.⁶⁹ Mixing was performed by gentle swirling using the pipette tip. A590/A800 values taken before DMSO addition were subtracted from values taken after DMSO addition to measure solubilized indigoidine. The resulting values were then normalized to the 0 μ M pigment condition. A concentration of 70 % DMSO with this absorbance measurement method could detect a broad range of pigment concentrations.

To measure the absorbance of indigoidine produced by cell-free expressed BpsA-TC, base reaction mixes with BAP1 lysate and pET28a(+)-BpsA-TC were incubated at 30°C.

Modifications to the reactions for validating pigment production are described in the text. All reactions were laid out on a 384-well plate and measured after ~ 20 h. A590 and A800 measurements were taken before and after 70 % DMSO addition and normalized as described above. Values reported for the L-Gln, CoA, and DNA experiments were further normalized to appropriate baseline conditions (0 mM L-Gln or CoA and 0 ng/ μ L DNA).

5.3.7. Comparative quantification of TC-based fluorescence measurements

The distribution of reactions across a plate and subsequent analyses with FIAsh-EDT₂ and ReAsH-EDT₂ labeling reagents (Thermo Fisher #T34563) were evaluated to correct for positional effects from the plate and reader²²⁷. An appropriate design, tested with protein standards, was then used in all experiments comparing endpoint TC/FIAsh measurements among reactions within a single plate read (e.g., different concentrations of PEP). For each condition tested (e.g., specific PEP concentration), a triplicate set of test reactions (with plasmid) and its corresponding triplicate set of control reactions (without plasmid) were prepared (10 μ L volumes). Triplicate test reactions were then plated adjacently on one row (e.g., Row H) and control reactions on another row (e.g., Row I). The test set and control set of each condition were laid out symmetrically to minimize well-to-well variation due to plate design.^{227,228} After reaction incubation, 0.5 μ L of the labeling reagent was added with gentle swirling of the pipette tip to each well in the triplicate set of test reactions then immediately to the corresponding control set. This was done for each condition before moving to the next. The plate was then read without a lid on a BioTek Cyation5 (Agilent) reader. For the experiment comparing labeling reagents, fluorescence measurements were taken with top optics every 5 minutes for almost an hour using appropriate filter sets (ReAsH: Exc 590 / 15 nm, Em 610 / 15 nm; FIAsh: Exc 510 / 15 nm, Em 530 / 15 nm) and autogain, with 2 s shaking at 567 cycles per minute linear frequency before every read. For every other experiment with FIAsh, measurements were taken after a 10 min incubation period at room temperature (Exc 510 / 15 nm, Em 530 / 15 nm, Gain 70). To isolate signals from

TC/FIAsH complexes in the data, signals from control reactions were subtracted from those of test reactions. To mitigate row effects from the detector (which were apparent in our data because our microplate reader reads row-wise) all resulting differences were normalized to the minimum value.²²⁷ Averages with standard error bars were plotted to visualize relative expression changes. For relative intensity analyses (%), the averaged values and standard errors were recalculated by multiplication to a constant of 100 over the maximum average value.

5.3.8. BpsA concentration estimation using TC/FIAsH measurements

Triplicate reactions containing pET28a(+)-BpsA-TC and triplicate controls with no DNA were set up on the same row in a microplate with reaction mixes (without DNA) laid out on an adjacent row. All reactions were prepared in 10 μ L volumes. After a ~ 20 h incubation at 25 °C, the blank reaction mixes on one row were spiked with 1 μ L of water or purified BpsA-TC protein dissolved in water to prepare triplicate sets of reactions containing 0, 50, 100, 200, 400 μ g/mL protein for standard curve generation. The test and control reactions on the adjacent row were spiked with 1 μ L water to equalize volumes. FIAsH (0.5 μ L) was added to each well and the plate was read following the protocol described above. Signals for reactions from either row were normalized to the minimal signal within the same row (i.e., lowest signal among the 3 control reactions). The appropriate dilution factor was applied. Linear regression using average normalized values from standard reactions was used to estimate the concentration of expressed BpsA-TC in the test reaction. This analysis was performed for each reaction mix (base vs optimized) when quantifying BpsA-TC since TC/FIAsH complex formation is likely to alter in different reaction backgrounds (**Appendix Figure 5.10**).

5.3.9. TycA and Pys concentration estimation

Triplicate reactions used to make comparative TC/FIAsh measurements between base and optimized conditions were pooled. Approximately 45 μ L of pooled lysate reactions were run through equilibrated Strep-Tactin®XT Spin Columns (IBA Life Sciences, #2-4150-025). Columns were washed 4x and eluted with 50 μ L elution buffer following the manufacturer's protocol. Elutions were loaded onto SDS-PAGE gels to visualize purified proteins.

5.3.10. SDS-PAGE gel and Western Blot analyses

For protein expressed *in vivo*, 5 μ L of purified BpsA-TC protein was denatured with 5 μ L 2x Laemmli sample buffer (BioRad #1610737). After boiling at 98°C for 7 minutes, 8 μ L of the denatured protein was loaded into a pre-cast 4-20% gel ordered from NuSep. The gel was run for 60 min at 250 V, fast stained (Fisher Scientific #2456) for 20 min, then de-stained in DI water overnight.

For protein expressed cell-free, triplicate reactions were pooled into a microcentrifuge tube immediately after FIAsh analysis. 5 μ L from the reaction was added to 5 μ L of a loading buffer mixture containing 20 mM DTT and 2x Laemmli sample buffer (BioRad #1610737) in a PCR tube. Tubes were incubated in boiling water for 5 minutes, then 2 μ L of each sample was loaded into 7.5% pre-cast gels (BioRad #4561026). Gels were run for 10 minutes at 80 V then 90 min at 250 V. After being stained (Thermo Fisher #LC6060) for 60 min, gels were de-stained in DI water overnight. This protocol was also used to visualize elutions of StrepII-tagged TycA and Pys, but μ L of sample was used to load SDS-PAGE gels.

For Western Blot analysis, a Bio-Rad Trans-Blot SD Semidry Transfer Cell was used to transfer bands onto a PVDF membrane. The membrane was probed with 6xHis-tag Monoclonal Antibodies (R&D Systems #MAB050) overnight after blocking with BSA in TBS Buffer (Thermo Fisher #37520). Blots were incubated with the relevant secondary antibody, HRP conjugate then developed with CN/DAB Substrate (Thermo Fisher #34000).

5.3.11. Quantification of active sfGFP

Fluorescence measurements of reactions expressing sfGFP were taken using top optics on a BioTek Cytation5 (Agilent). Excitation and emission filters were set to 485 nm and 538 nm, respectively.

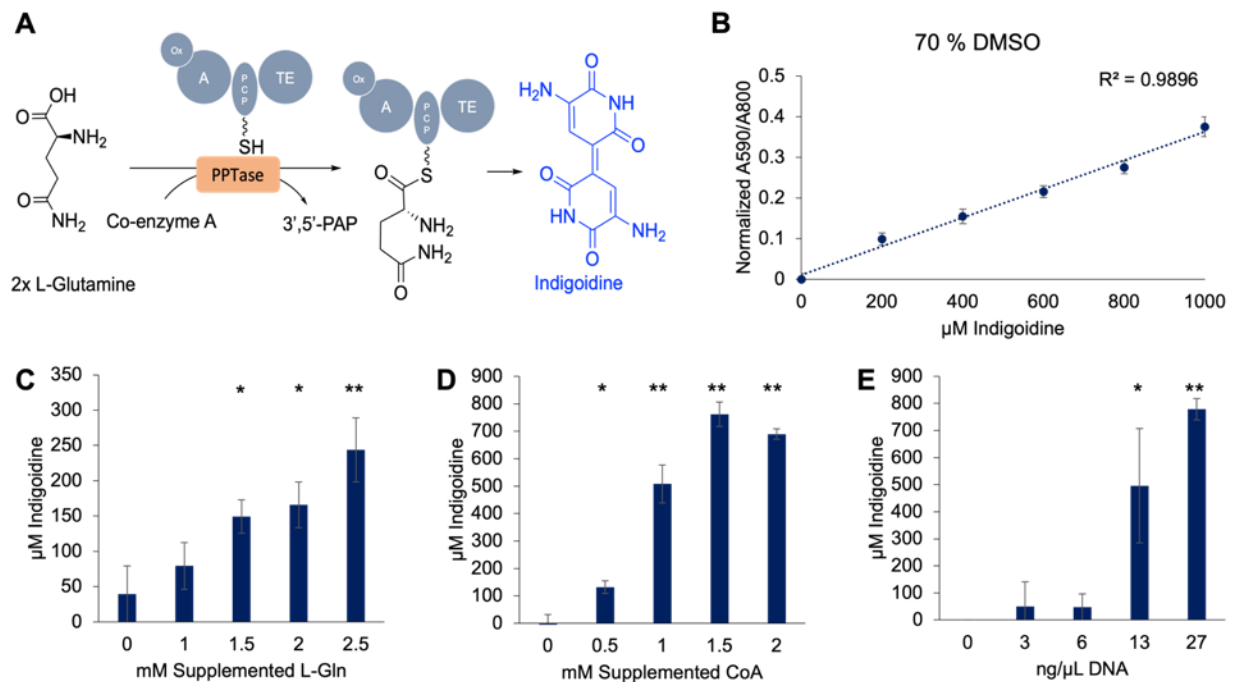
5.4. Results and Discussion

5.4.1. Cell-free expression of functional BpsA-TC in *E. coli* lysates

The activities of enzymes can affect their cell-free production if the enzymes' catalytic domains share substrates with the TX/TL machinery. To study this phenomenon for NRPS CFE using TC tag technology, it is important that the TC tag itself does not abolish the catalytic function of cell-free expressed NRPSs in *E. coli* lysates. To this end, a pET28a(+) vector containing the BpsA coding sequence and a C-terminal TC tag sequence to express TC-tagged BpsA (BpsA-TC) was constructed. BpsA is a great model for investigating NRPS CFE due to its relatively simple structure as a single module NRPS and the optical detectability of its product^{220–222}. BpsA synthesizes one molecule of indigoidine, a blue pigment that absorbs light at a wavelength of 590 nm, from two molecules of L-glutamine (L-Gln) (**Figure 5.1A**)⁶⁹. To do so, BpsA must be functionally activated into its *holo*-form by a 4'-phosphopantetheinyl transferase (PPTase) which phosphopantetheinylates the PCP domain. BAP1 is a BL21(DE3) derivative strain with a genomically integrated *sfp* gene that expresses a PPTase²¹⁴.

Detecting the blue pigment in BAP1 lysate reactions incubated with the plasmid construct would indicate successful expression of full-length and functional BpsA-TC protein. An assay was therefore established, allowing indigoidine measurement in a lysate-based cell-free system. To achieve this, purified indigoidine was spiked into lysate reactions without DNA. A sufficiently broad range of pigment concentrations could be detected by normalizing the absorbance of the pigment at 590 nm to lysate turbidity absorbance at 800 nm, and then coupling these normalized values

Figure 5.1. Indigoidine production via BpsA in CFE reactions. **A)** Schematic of BpsA-catalysed conversion of 2x L-Gln to 1x indigoidine. The *apo*-form of BpsA is activated by the addition of a coenzyme A and the presence of PPTase. The active *holo*-form BpsA converts two L-Gln molecules to indigoidine, involving oxidation (Ox), adenylation (A), peptide carrier protein (PCP), and thioesterase (TE) domains. **B)** Standard curve of indigoidine generated by spiking increasing concentrations of purified indigoidine into BAP1 lysate CFE mock reactions. Indigoidine in the lysate background was solubilized by adding 70 % DMSO directly to individual reactions. Absorbance measurements were taken at 590 nm (A590) and divided by signals at 800 nm (A800) to deconvolute pigment absorbance from lysate turbidity. Afterwards, A590/A800 signals before DMSO addition were then subtracted from measurements after 70 % DMSO addition to measure soluble pigment.⁴² For visualization of trends, values were averaged, compared to a baseline condition, and plotted with error bars representing the standard error of the mean (n=3). **C, D, E)** Validation of indigoidine production from functional BpsA-TC protein in CFE reactions using the indigoidine detection method established here. CFE reactions with 13 ng/μL plasmid were initiated with increasing concentrations of **C)** supplemented L-Gln (no CoA) or **D)** supplemented CoA (no L-Gln). **E)** CFE reactions containing 1 mM CoA were initiated with different concentrations of plasmid DNA. Asterisks represent significant differences compared to the no L-Gln, no CoA, or no DNA conditions, evaluated with a Student's t-test. * p-value < 0.05, ** p-value < 0.01.



to the addition of 70 % DMSO directly to the mock reactions (**Figure 5.1B**, **Appendix Figure. 5.7**). This optimized protocol was adapted from previous reports of indigoidine measurement in cell cultures and purified samples^{69,198}.

With an assay established, pigment production from the pET28a(+)-BpsA-TC construct could be tested in overnight CFE reactions. Reactions initiated with higher concentrations of the indigoidine precursor, L-Gln, have increasing normalized absorbance values, implying pigment production (**Figure 5.1C**). Separately, the base reaction mix was supplemented with CoA, from which the PPant post-translational modification that functionalizes the NRPS PCP domain is derived. Absorbance measurements also increase when up to 2 mM CoA was added to the reaction (**Figure 5.1D**). The improvement in pigment production in CoA-supplemented reactions is drastic even without supplemented L-Gln. To further validate that the measurements correspond to functional BpsA expressed from construct, 1 mM CoA was added to base reaction mixtures containing varying concentrations of the plasmid. Consistent with previous observations, the DNA template concentration affects protein expression and therefore product formation (**Figure 5.1E**). Altogether, the system expresses BpsA-TC and the terminal TC tag does not abolish BpsA folding and activity. Additionally, while optimizing enzyme turnover is outside the scope of this work, it is important to note for future efforts to advance NRPS CFE that these lysates generated a maximum concentration of ~ 900 μ M or ~ 223 mg/L indigoidine which is higher than indigoidine titers reported in the PURE cell-free system (250 μ M)²⁰⁷.

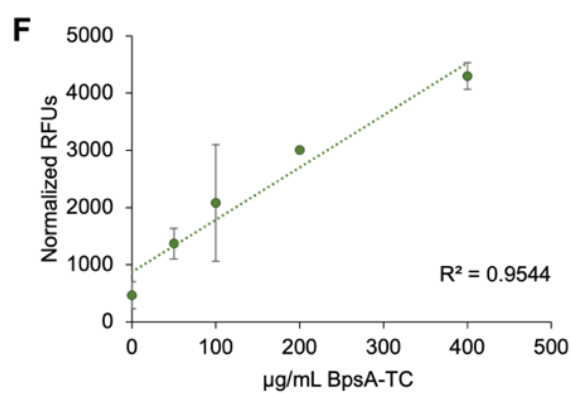
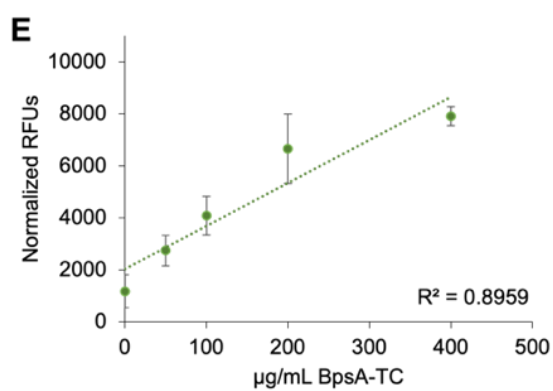
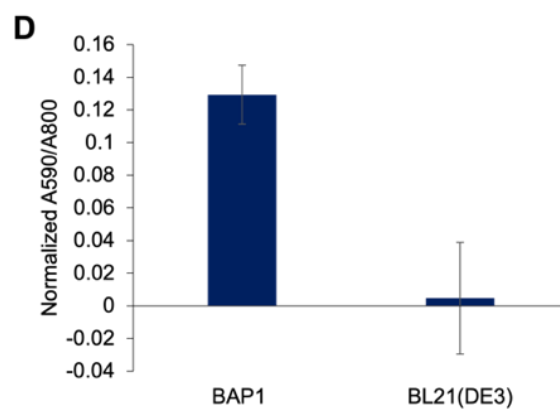
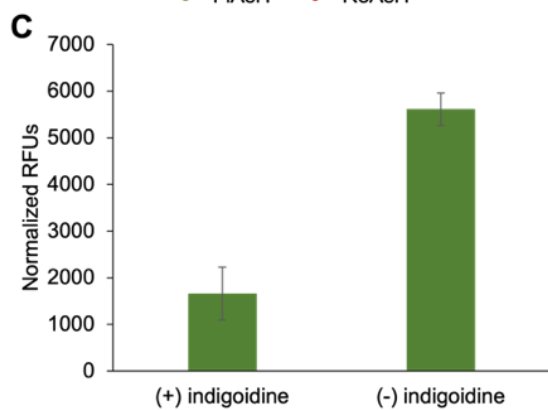
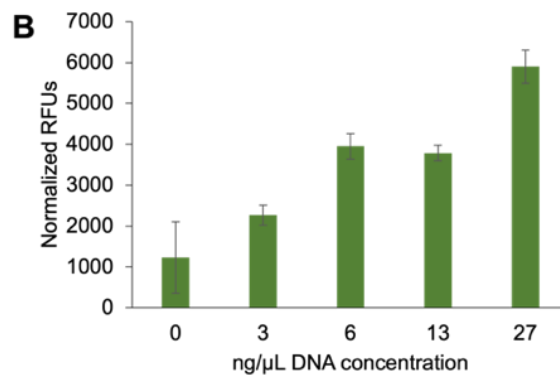
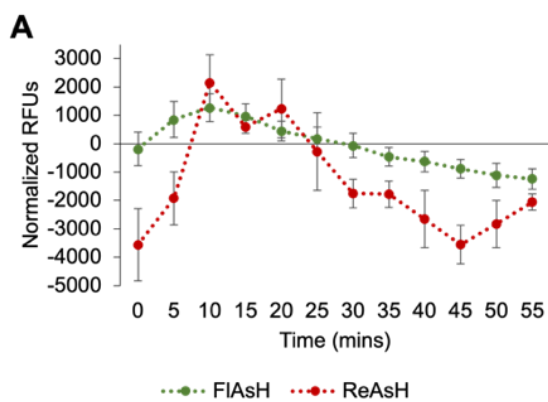
5.4.2. Optimization of a method for TC-based fluorescence measurement in an *E. coli* lysate CFE system

Given that a functional TC-tagged NRPS can be expressed in the system, a protocol for TC tag detection in lysate-based CFE reactions could be established. The TC tag, optimized for *in vivo* applications, has successfully been adapted for an *in vitro* expression study in the PURExpress system, where the tag was complexed with the organoarsenic ReAsH reagent to

emit a fluorescent signal at 610 nm²¹³. Reportedly, the TC/ReAsH complex has a low signal-to-noise ratio in lysate reaction backgrounds, but the use of controls to mitigate noise has not been explicitly stated in the literature²¹³. Alternatively, the organoarsenic FIASH reagent, which emits at 530 nm, has been used to detect TC-tagged proteins with higher specificity than ReAsH in cells but has not been tested in lysate reactions²²³.

In overnight reactions that expressed BpsA-TC, signals above background are highest after a 10-minute incubation period with the same concentration of either reagent (**Figure 5.2A**). This is sufficient time for the TC/fluorophore complex to form based on a previously described protocol in cells²²³. The signal in BpsA-TC expressing reactions however decreases at later timepoints, consistent with the manufacturer's descriptions of these reagents binding to the background. The FIASH reagent was therefore selected over ReAsH to obtain end-point measurements of BpsA-TC production since the former emitted less noisy and variable signals. To verify that the signal obtained at 10 minutes corresponds to synthesized BpsA-TC, FIASH was added to overnight BAP1 lysate-based reactions prepared with varying DNA concentrations. Evidently, increasing concentrations of the expression template in reactions results in higher fluorescence readouts, suggesting the detectability of BpsA-TC in lysate reactions with the FIASH reagent (**Figure 5.2B**). However, the signal does not increase between the 6 and 13 ng/μL DNA conditions which contradicts the trend in indigoidine production in the same lysate (**Figure 5.1E**). It was hypothesized that fluorescence signals from reactions that produce significant pigment quenched due to possible overlap in the indigoidine absorbance and TC/FIASH emission spectra. To evaluate this possibility, *apo*-form BpsA-TC was purified from *E. coli* BL21(DE3) cells lacking the *sfp* gene for PPTase expression. Purified BpsA-TC was then added into BL21(DE3) lysate-based reaction mixes with or without purified indigoidine, demonstrating that the presence of pigment does dampen TC/FIASH emissions (**Figure 5.2C**). This quenching issue should thus be considered when pairing the fluorescence protein labeling method with the detection of other enzymatic products with optical properties. Moving forward, the TC/FIASH method is further

Figure 5.2. Optimization of TC/FIAsH complex measurements in *E. coli* lysate-based CFE reactions. **A)** Analysis of FIAsH vs ReAsH utility for measuring purified BpsA-TC spiked into BAP1 lysates containing 1 mM CoA. 15 μ M of each organoarsenic reagent was added to overnight reactions started with pET28a(+)-BpsA-TC and their corresponding controls. Fluorescence measurements were taken every 5 minutes for 1 h and signals from reactions with cell-free expressed BpsA-TC were normalized to controls. **B)** End-point TC/FIAsH measurements in BAP1 lysate-based reactions incubated overnight with 1 mM CoA and increasing amounts of starting pET28a(+)-BpsA-TC DNA concentrations. Values are normalized RFUs taken after 10 min of reaction incubation with 15 μ M FIAsH. **C)** Signal interference due to TC/FIAsH fluorescence quenching by indigoidine. TC/FIAsH fluorescence measurements of purified BpsA-TC protein (100 μ g/mL) in mock lysate reaction mixes with or without 200 μ g/mL indigoidine. RFUs were taken after 10 min reaction incubation with 15 μ M FIAsH and normalized to respective no protein controls. **D)** Absorbance measurements of indigoidine in BAP1 and BL21(DE3) lysate-based reactions after overnight BpsA-TC expression, normalized to appropriate controls without DNA template. **E, F)** TC/FIAsH fluorescence standard curves generated by adding 30 μ M FIAsH to *E. coli* **E)** BL21(DE3) and **F)** BL21 Star (DE3) lysate reaction backgrounds carrying purified BpsA-TC standards. The dotted line in each graph represents the least squares regression. Reactions were run in triplicate. For all graphs, averaged normalized values are visualized with error bars representing the standard error of the mean (n=3).



optimized in lysates from the *E. coli* BL21(DE3) strain, which lack the PPTase and do not produce pigment after overnight BpsA-TC expression (**Figure 5.2D**), and in its derivative BL21 Star(DE3) strain, which should similarly express the *apo*-form of the enzyme.

In overnight reactions that expressed BpsA-TC, signals above background are highest after a 10-minute incubation period with the same concentration of either reagent (**Figure 5.2A**). This is sufficient time for the TC/fluorophore complex to form based on a previously described protocol in cells²²³. The signal in BpsA-TC expressing reactions however decreases at later timepoints, consistent with the manufacturer's descriptions of these reagents binding to the background. The FIAsH reagent was therefore selected over ReAsH to obtain end-point measurements of BpsA-TC production since the former emitted less noisy and variable signals. To verify that the signal obtained at 10 minutes corresponds to synthesized BpsA-TC, FIAsH was added to overnight BAP1 lysate-based reactions prepared with varying DNA concentrations. Evidently, increasing concentrations of the expression template in reactions results in higher fluorescence readouts, suggesting the detectability of BpsA-TC in lysate reactions with the FIAsH reagent (**Figure 5.2B**). However, the signal does not increase between the 6 and 13 ng/μL DNA conditions which contradicts the trend in indigoidine production in the same lysate (**Figure 5.1E**). It was hypothesized that fluorescence signals from reactions that produce significant pigment quenched due to possible overlap in the indigoidine absorbance and TC/FIAsH emission spectra. To evaluate this possibility, *apo*-form BpsA-TC was purified from *E. coli* BL21(DE3) cells lacking the *sfp* gene for PPTase expression. Purified BpsA-TC was then added into BL21(DE3) lysate-based reaction mixes with or without purified indigoidine, demonstrating that the presence of pigment does dampen TC/FIAsH emissions (**Figure 5.2C**). This quenching issue should thus be considered when pairing the fluorescence protein labeling method with the detection of other enzymatic products with optical properties. Moving forward, the TC/FIAsH method is further optimized in lysates from the *E. coli* BL21(DE3) strain, which lack the PPTase and do not produce

pigment after overnight BpsA-TC expression (**Figure 5.2D**), and in its derivative BL21 Star(DE3) strain, which should similarly express the *apo*-form of the enzyme.

To use TC/FIAsh signals for measuring protein CFE in lysate reactions, an optimal concentration of FIAsh for detecting a range of BpsA-TC concentrations in *E. coli* lysates should be established. To this end, BpsA-TC expressed and purified *in vivo* was used to generate standards for optimization of the TC/FIAsh method (**Appendix Figure 5.8**). Evidently, fluorescence signals for a range of protein concentrations (0 to 400 µg/mL) in BL21(DE3) lysates are most distinguishable with 30 µM FIAsh reagent compared against concentrations of 7.5 and 15 mM FIAsh, seemingly because these lower concentrations are insufficient to overcome background noise from the lysate (**Figure 5.2E, Appendix Figure 5.9**). 30 µM FIAsh is similarly useful in BL21 Star(DE3) lysate reactions (**Figure 5.2F**). Thus, TC/FIAsh measurements can be obtained when either lysate is used as an expression platform for BpsA. However, BpsA-TC can be best validated via SDS-PAGE when expressed in BL21 Star(DE3) lysate reactions incubated at 25 °C (**Appendix Figure 5.11**). Since SDS-PAGE can serve as an orthogonal method for validating correlations of TC/FIAsh signal trends to protein CFE, succeeding reactions were performed in these lysates at this temperature.

Having established parameters for detecting BpsA-TC in lysates using FIAsh-based and orthogonal SDS-PAGE methods, the TC/FIAsh complex's utility for measuring NRPS CFE under different reaction conditions could be evaluated. The large sizes, complex structures, catalytically active domains, and precursor (i.e., amino acids) and cofactor requirements (i.e., magnesium and ATP for adenylation) of NRPSs could constrain TX/TL processes and therefore their full-length synthesis, but evidence of such limitations and methods for probing these with high throughput have not been reported²¹². To address this, the ability of the optimized TC/FIAsh detection protocol to interrogate NRPS CFE is assessed using BpsA with an active A domain (a core component of NRPSs) as a model¹⁹⁰. This strategy is specifically applied here to determine whether resource competition problems exist for three crucial CFE reaction components that are

known to energize TX/TL and background lysate metabolism, but could also feed into the typical NRPS A domains—PEP, magnesium, and amino acids. However, a potential caveat of the current fluorescence labeling method is the influence of the physicochemical environment on TC/FIAsH complex formation. Indeed, drastically changing the composition of reactions by removing all or one of these three components affects TC/FIAsH signals (**Appendix Figure 5.11**). A possible explanation for this is that altering the ionic strength (by removing magnesium) and pH (by removing amino acids or PEP) of the system impacts binding of fluorophore's arsenic groups to the tag's cysteine groups. For this reason, the effect of each reagent on BpsA-TC CFE was evaluated separately and optimized sequentially. SDS-PAGE gels were crucial for validating a testable range of reaction concentrations wherein trends in fluorescence measurements correlate with TC-tagged protein. The results show that despite the method's caveats, expression trends can be determined from TC/FIAsH measurements for the reported ranges of reagent concentrations that are still relevant to lysate-based CFE.

5.4.3. Unique reaction conditions are required for the full-length synthesis of different functional proteins

5.4.3.1. PEP conditions

An obvious potential limitation to functional NRPS CFE is ATP availability. TX/TL processes that depend on ATP include dNTP synthesis, aminoacyl-tRNA charging, and GTP regeneration for peptide bond formation²⁰⁹. In the PANOX-SP system, ATP generation is sustained by ADP phosphorylation when the energy substrate PEP is converted to pyruvate. Without other limiting factors, a higher PEP requirement should theoretically better fuel the sufficient expression of a large protein (> 100 kDa) compared to smaller model CFE proteins (e.g. CAT and sfGFP) which are < 30 kDa²⁰⁹. The degradation of PEP itself however contributes to inorganic phosphate accumulation which in turn inhibits cell-free protein synthesis²²⁹. Furthermore, ATP overproduction negatively impacts CFE efficiency²¹⁰. Thus, for any given

protein, an optimal range of initial reaction PEP concentrations exists where PEP energizes but is not inhibitory towards TX/TL^{229–231}. This range is further defined by central metabolic reactions that consume ATP in an unproductive manner (i.e., not contributing to protein expression)^{74,154,210}. In the case of NRPS CFE, the A domain that is characteristic of these enzymes could also deplete the ATP pool, further contributing to expression limitations^{190,204}. The TC/FIAsH protocol was thus tested as a potential strategy for assessing the sensitivity of BpsA expression to PEP concentrations, and for investigating whether the catalytic A domain also competes for this TX/TL resource.

Fluorescence reads of the BpsA-TC reactions supplemented with FIAsH reagent show a trend in signals that were corroborated by an SDS-PAGE analysis, demonstrating the method's applicability for measuring relative expression in reactions fed different concentrations of PEP. To interrogate the interdependency between NRPS TX/TL and NRPS catalytic activity, a construct that would express BpsA-TC with a glutamine-to-alanine substitution at a catalytically essential residue in the active site of the A domain was generated. This point mutation (E315A) disrupts magnesium ion binding which stabilizes ATP in the active site (**Appendix Figure 5.12**)²³². The mutant protein, BpsA-TC_E315A, would thus have an inactive A domain due to disrupted ATP binding. Where BpsA-TC synthesis in this system peaks with higher starting PEP concentrations (55-77 mM), the maximal BpsA-TC_E315A expression already occurs in the reaction initiated with 33 mM PEP (**Figures 5.3A & B**). Trends for BpsA-TC expression under the tested PEP concentrations can be corroborated by SDS-PAGE gels (**Appendix Figure 5.13A**). These results confirm that functional NRPS TX/TL can have higher energy substrate requirements than proteins of the same size because the catalytically active A domain contributes to the additional catabolism of ATP. Indeed, BpsA-TC_E315A is expressed more than BpsA-TC in reactions that are read at the same time and that were initiated with the same amount of PEP.

Reactions were also set up with sfGFP to compare the BpsA trends to those of a smaller, robustly folding protein (**Figure 5.3C**). Interestingly, relative expression patterns for sfGFP

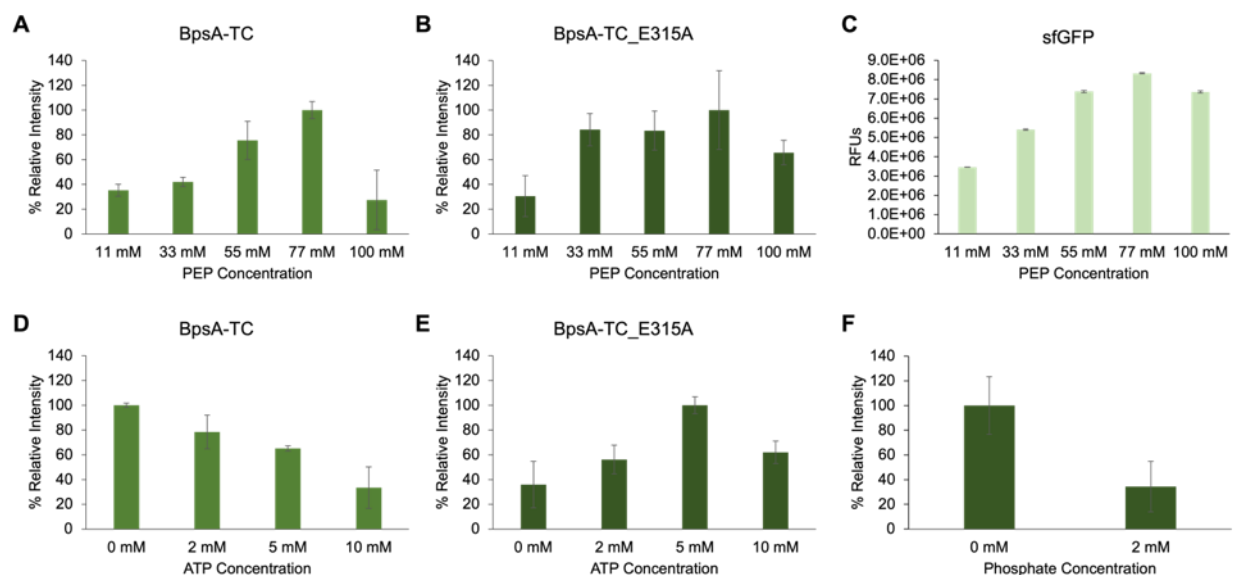


Figure 5.3. TC/FIAsh analysis of energy substrate requirements in NRPS CFE using BpsA as a model. **A, B, C)** Reactions were initiated with varying amounts of PEP to determine cell-free expression changes for **A)** BpsA-TC, **B)** BpsA-TC_E315A, and **C)** sfGFP. **D, E)** The tolerance of BpsA with or without an active A domain to ATP was similarly analyzed. Reactions with 77 mM PEP plus varying concentrations of ATP were started with plasmid expressing **D)** BpsA-TC and **E)** BpsA-TC_E315A. **F)** Reactions expressing BpsA-TC_E315A were initiated with or without 2 mM phosphate solution. All reactions were prepared with *E. coli* BL21 (Star)DE3 lysates. Reactions were run in triplicate and TC/FIAsh measurements were obtained by adding 30 μ M FIAsh reagent to each reaction after ~ 20 h incubation. TC/FIAsh measurements are presented as % relative intensities for visual comparison of trends between BpsA-TC and BpsA-TC_E315A data since these experiments were run on separate days. Error bars represent the standard error of the mean (n=3).

resemble those of BpsA-TC. However, sfGFP yields significantly increase as initial PEP concentrations rise to 77 mM while BpsA-TC_E315A expression has no significant change beyond the 33 mM PEP condition. Both BpsA-TC_E315A and sfGFP should not compete with TX/TL processes for ATP, so the disparity in these trends may be attributed to other factors that limit the full-length synthesis of a larger and likely less soluble protein than sfGFP.

Unexpectedly, BpsA-TC synthesis in reactions initiated with 100 mM PEP is substantially reduced which may also be attributed to the active A domain (**Figure 5.3A**). In the conventional PANOx-SP system, PEP concentrations beyond the optimum cause high inorganic phosphate accumulation due to the degradation of PEP itself²²⁹. In the current system, this phenomenon may be partially responsible for the significant inhibition of BpsA-TC synthesis in the 100 mM PEP reactions compared to all other conditions. However, BpsA-TC_E315A yields in reactions started with 100 mM PEP are not as drastically reduced (**Figure 5.3B**). These data imply that the A domain contributes to the inhibition of NRPS TX/TL when PEP is increasingly available, hypothetically because there is a larger ATP pool to liberate pyrophosphates from during amino acid adenylation^{190,229}. To test whether the active A domain indeed leads to lower expression of BpsA when more ATP is available, reactions containing constructs for BpsA-TC or BpsA-TC_E315A were fed 77 mM PEP and increasing supplemental concentrations of ATP. PEP was still added to the system because adding ATP alone resulted in no protein expression. Compared to reactions without additional ATP, increasing ATP concentrations only reduces BpsA-TC expression (**Figure 5.3D**, **Appendix Figure 5.13B**) but improves or has no significant effect on BpsA-TC_E315A synthesis (**Figure 5.3E**). However, BpsA-TC_E315A synthesis decreases when reactions are supplemented with inorganic phosphates (**Figure 5.3F**). Altogether these data imply that the active A domain's liberation of free phosphates is increasingly deleterious to expression when higher ATP pools are available.

5.4.3.2. Magnesium conditions

Magnesium ions are essential in almost all stages of TX/TL. Among a multitude of roles, they stabilize mRNA molecules and ribosomes, are required in the aminoacylation of tRNAs, regulate the ionic strength of the lysate reaction environment, and serve as a co-factor for several TX/TL relevant enzymes^{233,234}. Particularly in cell extracts, where magnesium is also sequestered by enzymes of central metabolism, the magnesium requirement for maximum synthesis can vary per protein and even per batch of lysate^{209,229,235}. Methods for routinely optimizing the magnesium concentration of a specific target protein in a lysate-based cell-free system are thus necessary. Such protocols would be particularly beneficial in NRPS CFE efforts, given that enzymes belonging to this structurally diverse family may have varying magnesium concentration optima for TX/TL based on size differences alone^{229,235}. Additionally, NRPSs themselves use magnesium at their A domains to stabilize ATP binding when adenylating an amino acid substrate.^{190,204,232} Thus, along with metabolic processes that are counterproductive to TX/TL, this domain could also affect magnesium homeostasis in NRPS CFE reactions if it were to sequester free magnesium during *in vitro* synthesis.

If applicable, the TC tag and FIAsh fluorophore can be leveraged to understand and meet the magnesium requirements for the functional expression of an NRPS. BpsA-TC reactions were therefore prepared with varying concentrations of magnesium and 77 mM PEP since this concentration allowed the highest BpsA expression. The results, confirmed by SDS-PAGE, show that an optimal magnesium concentration for expression of the wild-type enzyme can be defined by the approach (**Figure 5.4A, Appendix Figure 5.14**). The expression trend for these reactions differ from those of BpsA-TC_E315A and sfGFP in the same reaction backgrounds, showing that the active A domain impacts NRPS CFE magnesium requirements and reinforcing that the optimal concentration of this TX/TL resource can vary from protein to protein (**Figures 5.4B & C**).²³⁵ Compared to sfGFP and BpsA-TC_E315A, BpsA-TC requires more magnesium for peak

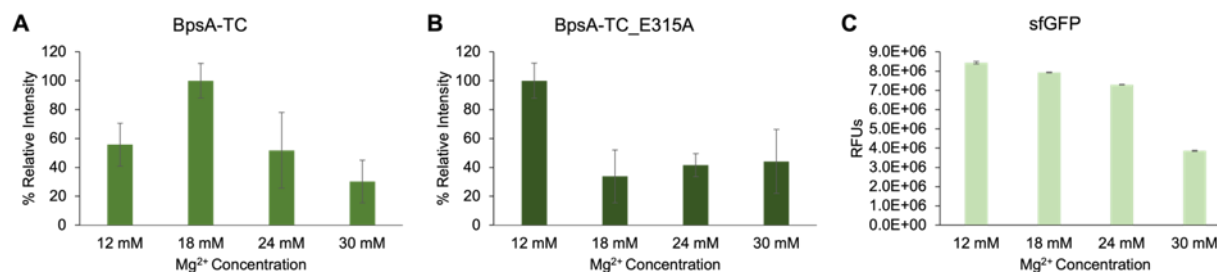


Figure 5.4. TC/FIAsH analysis of magnesium requirements in NRPS CFE using BpsA as a model. Reactions were initiated with varying amounts of Mg^{2+} to determine cell-free expression changes for **A)** BpsA-TC, **B)** BpsA-TC_E315A, **C)** sfGFP. All reactions were prepared with *E. coli* BL21 (Star)DE3 lysates. Reactions were run in triplicate and TC/FIAsH measurements were obtained by adding 30 μ M FIAsH reagent to each reaction after ~ 20 h incubation. TC/FIAsH measurements are presented as % relative intensities for visual comparison of trends between BpsA-TC and BpsA-TC_E315A data since these experiments were run on separate days. Error bars represent the standard error of the mean (n=3).

expression in the current system. A likely explanation is that additional phosphates liberated by the active A domain sequester free magnesium²³⁶.

5.4.3.3. Amino acids

Amino acids are the building blocks of peptides and are thus added into CFE reaction mixes to supply protein synthesis²³⁷. Like PEP and magnesium, this reaction component has been optimized for CFE in the *E. coli* PANOX-SP system, which is commonly initially supplemented with 2 mM of the 20 essential amino acids^{237–239}. This amount reportedly increases yields of CAT compared to lower or higher concentrations and has been previously adapted for NRPS expression in *E. coli* lysates^{54,211,239}. However, the sensitivity of NRPS expression to a range of amino acid concentrations should be assessed. As an NRPS is synthesized, its A domain(s) may recognize and adenylate certain amino acids that would otherwise be used for TX/TL^{190,199}. Additionally, NRPSs can have sequences with diverse amino acid distributions or A domains with different amino acid specificities¹⁹³. Also, compared to sfGFP, the CFE of a much larger protein like an NRPS would hypothetically require more free amino acids. Considering these, TC/FIAsh measurements can be used to probe NRPS production under different amino acid concentrations.

The initial concentrations of the amino acid mix (containing all 20 essential amino acids) in BpsA-TC expressing reactions were varied. Reactions were prepared with PEP and magnesium concentrations optimized for BpsA expression thus far (77 mM PEP and 18 mM magnesium). TC/FIAsh measurements from these reactions demonstrate no significant differences in expression levels between the 2, 4, and 6 mM conditions (**Figure 5.5A**). However, a substantial decline occurs when 8 mM of the amino acid mixture is used. These results are consistent with SDS-PAGE data (**Appendix Figure 5.15A**). On the other hand, BpsA-TC_E315A synthesis shows no significant change across all conditions, implicating the A domain's role in impairing BpsA expression presumably because of its interaction with L-Gln (**Figure 5.5B**)²⁴⁰. To investigate this, reactions containing 2 mM total amino acids and the constructs for BpsA-TC or

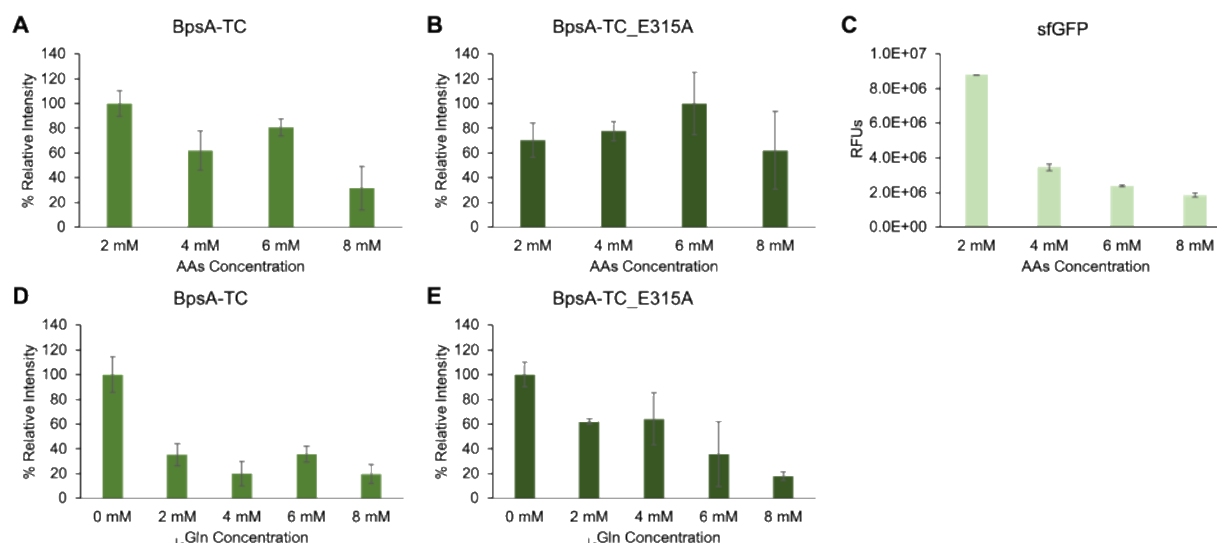


Figure 5.5. TC/FIAsH analysis of amino acid (AA) requirements in NRPS CFE using BpsA as a model. **A, B, C)** Reactions were initiated with varying amounts of the AA mix containing all 20 essential AAs to determine cell-free expression changes for **A)** BpsA-TC, **B)** BpsA-TC_E315A, **C)** sfGFP. **D, E)** BpsA, with or without an active A domain, expression with extra L-Gln was similarly analyzed. Reactions with 2 mM AA mix plus varying concentrations of L-Gln were started with plasmid expressing **D)** BpsA-TC and **E)** BpsA-TC_E315A. All reactions were prepared with *E. coli* BL21 (Star)DE3 lysates. Reactions were run in triplicate and TC/FIAsH measurements were obtained by adding 30 μ M FIAsH reagent to each reaction after ~ 20 h incubation. TC/FIAsH measurements are presented as % relative intensities for visual comparison of trends between BpsA-TC and BpsA-TC_E315A data since these experiments were run on separate days. Error bars represent the standard error of the mean (n=3).

BpsA-TC_E315A were also initiated with increasing concentrations of L-Gln. However, extra L-Gln in the reaction impairs expression of both proteins, but the mutant BpsA has a more gradual decline with increasing L-Gln supplement (**Figure 5.5D & E, Appendix Figure 5.15B**). Altogether, these findings are contrary to the assumption that reactions expressing BpsA with A domain mediated amino acid adenylating activity would benefit from more amino acids than the mutated counterpart. Expression trends for BpsA-TC and BpsA-TC_E315A also contradict the expectation that yields of larger proteins will increase with amino acid concentrations. Where sfGFP synthesis peaks at 2 mM, which was expected of this small protein in the PANOX-SP system, amino acid supplementation has no positive effect on the expression of both the wild-type and mutant BpsA (**Figures 5.5A-C**). Amino acid oxidation, which has been observed during lysate-based CFE, could be partially responsible for these observations²³⁶. Reactions that actively deplete amino acids like glutamate form phosphates²⁴¹. Amino acids also serve as secondary energy substrates that can drive ATP generation²³⁶, which in turn can also release phosphates. The deleterious effect of the free phosphate pool on BpsA expression is exacerbated by the A domain's participation in phosphate liberation (**Figures 5.3C-E**), which may explain the higher sensitivity of wild-type BpsA to amino acid concentrations compared to its mutant counterpart (**Figures 5.5A & B**). Further, the oxidation of certain amino acids in lysates reportedly occurs quickly, which may prevent the prolonged translation of larger enzymes that require these amino acids in batch reactions. For example, compared to one mole of sfGFP, BpsA requires 9x and 11x more moles of arginine and tryptophan, respectively, and both of these amino acids are depleted within 30 minutes in lysates^{236,241}.

In summary, BpsA expression was measured under different CFE reaction conditions to elucidate the sensitivity of this process to PEP, magnesium, and amino acid concentrations given the protein's physicochemical properties. The A domain's activity and the enzyme's structure were both investigated by comparing BpsA to BpsA-TC_E315A and BpsA-TC_E315A to sfGFP expression, respectively. Compared to standard CFE conditions, higher PEP and magnesium

concentrations benefit BpsA expression, largely due to the active A domain. As a result, PEP and magnesium concentrations that improve BpsA CFE are defined. Moving forward, it was determined whether these conditions that increase full-length BpsA synthesis would lead to a concordant increase in indigoidine production or similarly improve the expression of other monomodular NRPSs.

5.4.4. Nonribosomal peptide production improves with optimized NRPS expression

NRPSs are valuable for their bioactive peptide products. The previous sets of experiments show that full-length NRPS expression can be optimized by testing different CFE conditions with a C-terminal TC tag and the FIAsh dye. Higher concentrations of the fully synthesized enzyme should enable higher nonribosomal peptide production as well. To assess this, both BpsA-TC and indigoidine concentrations were estimated and compared in reactions made with base (containing 33 mM PEP and 12 mM magnesium) and optimized (containing 77 mM PEP and 18 mM magnesium) reaction mixes. Using TC/FIAsh measurements and appropriate standard curves, the average concentrations of BpsA-TC in reactions prepared with the optimized mix and base mix are estimated to be 95 $\mu\text{g/mL}$ and 56 $\mu\text{g/mL}$, respectively (**Figure 5.6A**). Reactions expressing BpsA-TC prepared with either mix were additionally initiated with purified PPTase, CoA, and L-Gln then incubated overnight. Using the colorimetric assay, titers of indigoidine were calculated from a standard curve and show a similar improvement after PEP and magnesium concentration optimization (160 μM and 99 μM indigoidine in the optimized and base reactions, respectively) (**Figure 5.6B**). As expected, optimizing CFE facilitates increased nonribosomal peptide production. Notably, the indigoidine titer in the optimized BL21 Star(DE3) lysate is lower than the maximum concentration obtained in the BL21(DE3)-derived BAP1 lysate ($\sim 900 \mu\text{M}$) (**Figure 5.1D**), despite BL21(DE3) being less productive for NRPS CFE compared to BL21 (Star)DE3 (**Appendix Figure 5.10A**)²⁰⁸. PPTase, CoA, and L-Gln concentrations can therefore be optimized to enable higher NRPS turnover in BL21 Star(DE3) lysates. Alternatively, a BL21

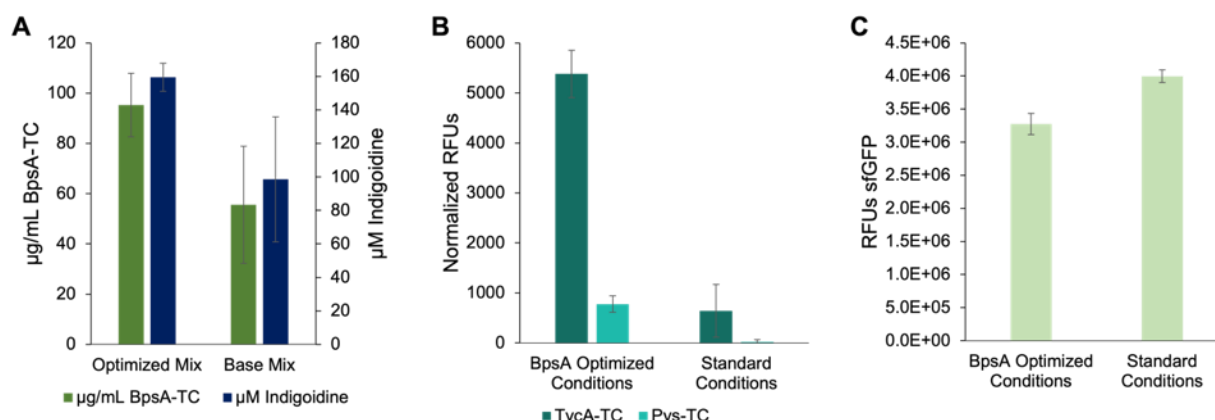


Figure 5.6. Conditions optimized for BpsA improve indigoidine production and the expression of similar NRPSs in lysates. A) PEP and magnesium optimization improve full-length, functional NRPS expression. BpsA-TC concentrations in reactions prepared with 77 mM PEP and 18 mM Mg^{2+} (optimized mix) or 33 mM PEP and 12 mM Mg^{2+} (base mix) were estimated with TC/FIAsH fluorescence standard curves for either reaction background. Indigoidine concentrations in lysate reactions incubated with 50 $\mu\text{g/mL}$ purified PPTase, 1 mM CoA, and 4 mM L-Gln in optimized or base reaction mixes were estimated using pigment standards. **B)** Compared to standard CFE conditions used in previous NRPS CFE studies (30°C incubation and base mix), the total set of conditions optimized for BpsA (25°C incubation and optimized mix) benefits the production of other monomodular NRPSs, TycA and Pys. **C)** Super folder GFP (sfGFP) lysate-based CFE using the condition set optimized for BpsA expression vs the standard condition set for protein synthesis in *E. coli* lysates. All reactions were run in triplicate. TC/FIAsH measurements were obtained by adding 30 μM FIAsH reagent to each reaction after ~ 20 h incubation. Averaged values are visualized with error bars representing the standard error of the mean. All reactions were prepared with *E. coli* BL21 (Star)DE3 lysates.

(Star)DE3 strain with a genome-integrated *sfp* gene could benefit the CFE of NRPSs, especially those that catalyze high-value but cytotoxic metabolites.

5.4.5. Conditions optimized for BpsA allow or improve the CFE of other single-module NRPSs

Standard CFE conditions commonly used in previous reports of lysate-based NRPS or NRPS-like protein expression include the preparation of reactions using the described base mix and their incubation at 30 °C^{54,211,212}. In theory, the CFE of NRPSs with similar architectures to BpsA will benefit from the total set of reaction conditions optimized here (i.e., optimized mix and 25 °C incubation). Other known monomodular NRPSs include tyrocidine synthetase I, TycA (124 kDa), which incorporates and isomerizes an L-phenylalanine residue into the cyclic antibiotic decapeptide tyrocidine²⁴², and pyreudione synthetase, Pys (142 kDa), which adenylates L-proline in the synthesis of amoebicidal pyreudiones²⁴³. In addition to being close in size to BpsA, these NRPSs each have an active A domain. TC- and StrepII-tagged versions of these NRPSs were expressed using standard CFE conditions and conditions optimized here for BpsA. TC/FIAsh measurements for both TycA-TC and Pys-TC show that these proteins are better synthesized under conditions optimized for BpsA expression. TycA expressed particularly well in the optimized system and could be visualized upon purification (**Appendix Figure 5.16**). Like TycA synthesized in base conditions, Pys produced in either condition set could not be observed in gels of lysate reactions or StrepII-tag elutions. On the other hand, the standard condition set, which was initially optimized for sfGFP²³¹, is still more beneficial for the synthesis of this protein compared to conditions optimized here for BpsA (**Figure. 5.6C**)²³¹. This implies that sfGFP is not universally informative when tailoring extracts for CFE, and further reinforces the need for reporters like the TC/FIAsh system when developing lysate-based platforms for different types of proteins. Together, these results confirm that CFE conditions optimized for a particular enzyme allow or improve the synthesis of proteins with similar architectures. Thus, while conditions optimized for

sfGFP production may not be universally optimal for larger biosynthetic enzymes, generalizable reporter tools like TC/FIAsH may be leveraged to construct model systems for studying and defining better CFE conditions of whole enzyme classes with high throughput. Notably, discrepancies in how much these conditions improved TycA vs Pys or BpsA expression suggest that TC/FIAsH measurements will still be useful for further probing and optimizing CFE on a protein-by-protein basis.

5.5. Conclusions

Lysate-based CFE is an accessible and scalable approach for expediting the expression of proteins that are difficult to synthesize *in vivo*, such as natural product forming NRPSs^{28,35,165}. Lysates may additionally be derived from strains that can post-translationally modify NRPSs into their active *holo*-forms^{212,214}, or microbial species that are already heavy producers of NRPSs^{215,217–219}. However, lysates possess complex interactions between the TX/TL machinery and metabolic reactions in the lysate that are unproductive to protein expression²¹⁰. The latter may include the catalytic activities of the enzyme product itself. Establishing the utility of protein fluorescence labeling in cell extracts, in addition to minimal expression systems²¹³, advances the investigation of these confounding relationships. It also provides insight into how protein size influences CFE requirements. These in turn enable the optimization of cell extracts for NRPS CFE. Here, a fluorescence-based protein detection method that employs a TC tag and FIAsH fluorophore was adapted for measuring NRPS expression in lysate-based CFE systems. This strategy can validate full-length NRPS translation in the lysate background with higher throughput than current NRPS detection methods. The approach allows the quantification of NRPS expression across several reactions with the use of careful controls that account for caveats of the dye in different reaction environments. Importantly, these measurements enable reaction optimizations that apply to the expression of other NRPSs as well. Overall, the method enables

rapid hypothesis generation and testing for broad applications in lysate-based NRPS CFE that could be extendable to other proteins. As demonstrated here with BpsA and its pigment product, this approach can also be complemented with an assay for indirectly measuring improvements to functional protein folding and post-translational modification²²⁶.

As demonstrated in this investigation, the method can be applied towards understanding requirements for full-length NRPS expression considering the mechanistic features of these enzymes (such as the role of ATP in catalysis). The conditions for the cell-free TX/TL of NRPSs are expected to change depending on the size, structure, and activity of the protein product itself. An objective here was to demonstrate that TC/FIAsH experiments can elucidate these relationships which previously lacked evidence. Comparisons of BpsA-TC, BpsA-TC_E315A, and sfGFP expression under various reaction conditions revealed the unique needs of NRPS CFE. This TC/FIAsH detection protocol may also be modified to meet specific experimental objectives under the umbrella of understanding NRPS CFE. For example, the TC tag may be appended to truncated forms of the NRPS coding sequence to quantify the frequency of premature NRPS translation. Or alternately, the TC tag could be appended N-terminally to demonstrate the degree of initiation of translation.

Measuring TC/FIAsH is also a powerful strategy for tailoring lysate-based reaction environments towards high-yield NRPS CFE. Throughout this investigation for example, the TC/FIAsH method identified better temperature, PEP, and magnesium conditions for monomodular NRPS CFE while providing insight into potential bottlenecks of productive amino acid utilization. The results of these experiments lend to further improvements in NRPS CFE systems. For example, the drastic inhibition of NRPS synthesis beyond a threshold concentration of PEP may be remedied using alternative ATP recycling schemes to prevent excessive free phosphate accumulation^{229,244}. Factors limiting productive amino acid utilization in NRPS CFE systems can also be further investigated. Apart from amino acid oxidation, it is also possible that concentrations of other reaction components such as tRNAs or aminoacyl-tRNA synthetases may

be insufficient for full-length large protein synthesis even when amino acid levels are high enough²³⁷. Moreover, extensive efforts to optimize lysates for high-yield production of other NRPS architectures would benefit from TC/FIAsH. While the results reported here are informative for monomodular NRPS CFE, the NRPS family of proteins is known to have high structural diversity¹⁹⁰. NRPSs with different sizes and numbers of A domains would therefore likely have unique optimal reaction condition sets. The TC/FIAsH method would be useful for defining these reaction parameters through sequential optimization experiments, in a way that offers more throughput than current methods to detect these enzymes^{54,211,212}.

Additional applications of the TC/FIAsH approach for NRPS CFE include discovery and prototyping. Despite their natural abundance, difficulty expressing NRPSs in *in vivo* systems has been a barrier to the characterization of these enzymes for natural product discovery²¹¹. The strategy could thus be usefully applied towards expressing libraries of NRPS sequences incorporating TC tags or prototyping genetic parts for successful NRPS expression in model or non-model lysate backgrounds. Measuring TC/FIAsH could also have value in screening engineered libraries, facilitating the discovery of structurally diverse nonribosomal peptides²⁴⁵.

Looking forward, this approach is envisioned to help expedite the study and optimization of lysate-based CFE systems for specific protein synthesis, especially enzymes like NRPSs that possess complex physicochemical properties which may impede cell-free TX/TL. This will consequently advance the development of robust NRPS expression platforms that will support academic and commercial applications.

Appendix

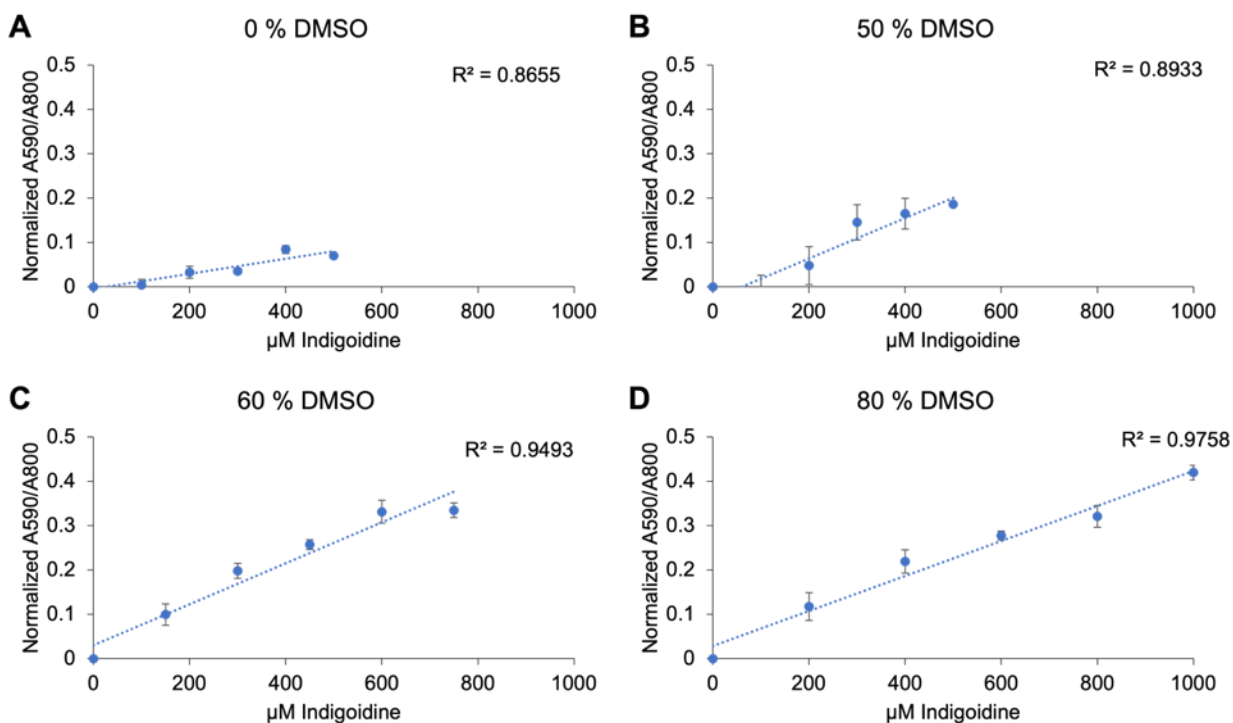


Figure 5.7. Standard curves of indigoidine from purified indigoidine in BAP1 lysate mock reactions. Due to the solubility limits of indigoidine, different concentrations of DMSO were added to the lysate reaction mix to dissolve indigoidine: **A)** 0% DMSO, **B)** 50% DMSO, **C)** 60% DMSO, **D)** 80% DMSO. In all cases, absorbance measurements were taken at 590 nm and divided by signals at 800 nm to deconvolute pigment absorbance from lysate turbidity.¹⁹⁸ Afterwards, A590/A800 signals before DMSO addition were then subtracted from measurements after 70 % DMSO addition to measure soluble pigment.⁶⁹ For visualization of trends, values were averaged, compared to the baseline condition (0 μM indigoidine), and plotted with error bars representing the standard error of the mean (n=3).

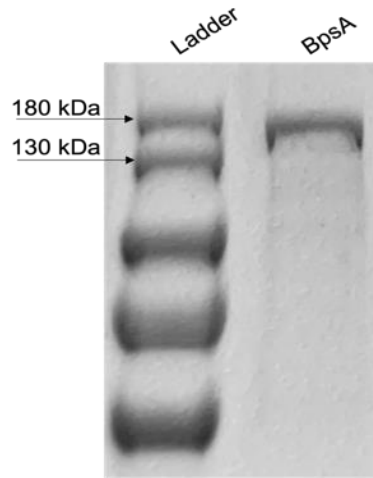


Figure 5.8. SDS-PAGE gel of purified BpsA-TC expressed heterologously in *E. coli* BAP1.

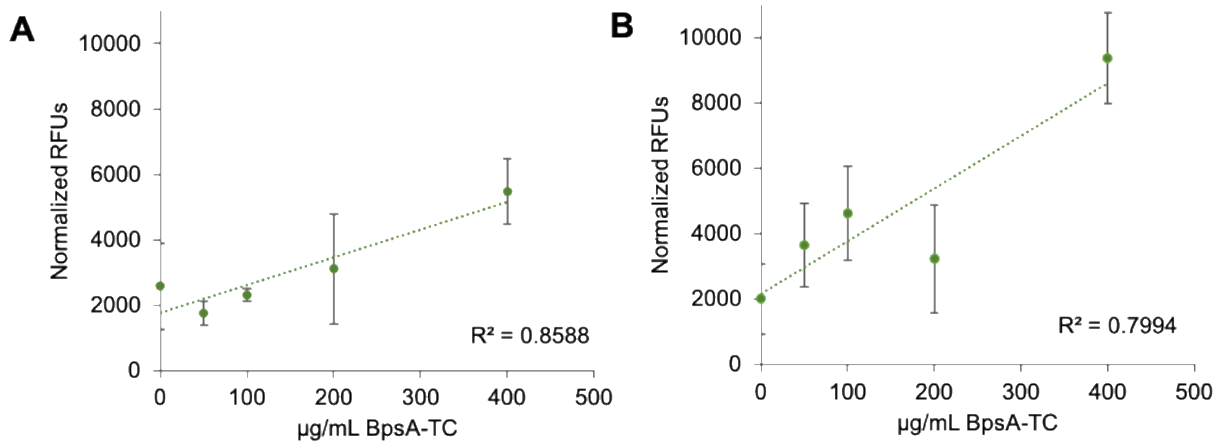


Figure 5.9. Optimization of FIAsH concentration for TC/FIAsH measurement and validation in *E. coli* lysate-based reactions. Increasing concentrations of purified BpsA-TC were spiked into BL21 (DE3) lysate-based CFE mock reactions. 0.5 µL FIAsH was added to different final concentrations for each set of reactions to determine an effective FIAsH concentration for measuring a range of TC-tagged protein in lysates: **A)** 7.5 µM, **B)** 15 µM. Averaged normalized values are visualized with error bars representing the standard error of the mean (n=3).

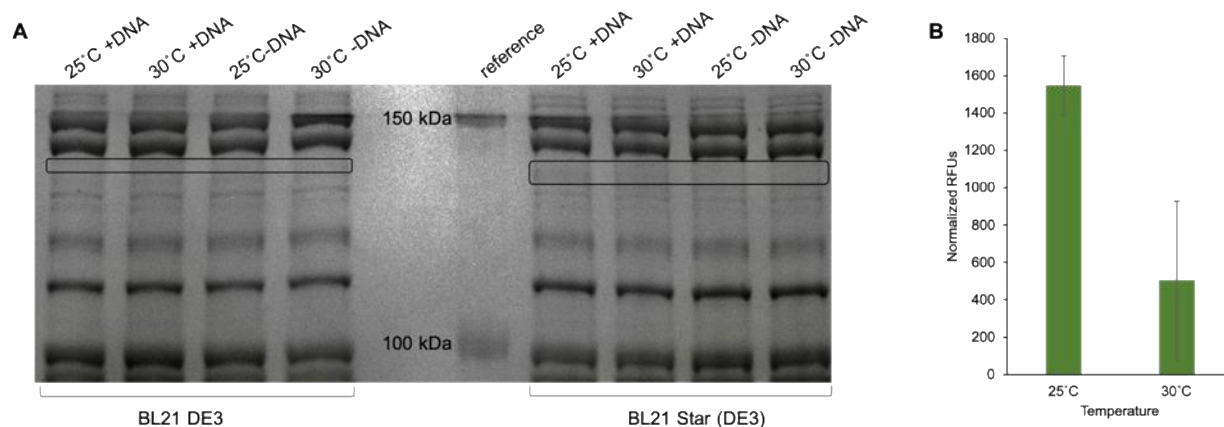


Figure 5.10. BpsA-TC CFE in different temperatures and *E. coli* lysates. A) SDS-PAGE gel of CFE protein at 25°C and 30°C of BL21(DE3) lysate and BL21 Star (DE3) lysate using the CFE base conditions. Lanes represent pooled reactions (n=3). A faint band is most visible compared against the control for BL21 Star (DE3) reactions started with DNA and incubated at 25°C. Difficulty expressing other NRPSs in BL21(DE3) lysates has been reported and is presumably due to the instability of NRPS mRNA in this system²¹¹. **B)** Comparison of BpsA-TC synthesized in BL21 Star (DE3) lysate reactions under different incubation temperatures (25°C vs 30°C), analyzed with 30 µM FIAsH. These results reinforce higher expression of BpsA-TC in BL21 (Star) DE3 lysate reactions incubated at 25 °C, compared to 30 °C, which is consistent with reports of higher soluble BpsA expression in cells at lower temperatures^{226,246}. Reactions were run in triplicate and TC/FIAsH measurements were obtained by adding 30 µM FIAsH reagent to each reaction after ~ 20 h incubation. Averaged normalized values are visualized with error bars representing the standard error of the mean (n=3).

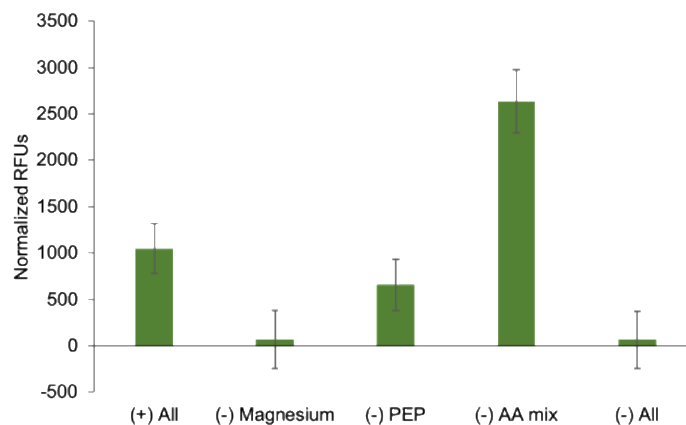


Figure 5.11. TC/FIASH measurements in different reaction backgrounds. Magnesium, PEP, the amino acid (AA) mix, or all of these components were removed from mock reaction mixes containing 100 $\mu\text{g/mL}$ purified BpsA-TC. Appropriate controls were prepared, and all samples were distributed on a 384-well microplate. The mixes were incubated at room temperature for 1 h. Afterwards, 30 μM FIASH reagent was added to each sample and all wells were read with the appropriate filter set after 10 min. Averaged normalized values are visualized with error bars representing the standard error of the mean ($n=3$).

difD	ITYGTEPMPQSTLRGIHRLFPDVRLLKQTYGLTEL	GIFSTKSKD----	SQSTWMKVGG----	287
hasG	VTYGTEVMPESVLSNLNQLFPNIRFHQTYGLTEL	GIMRSKSR-----	SDSLWFKVGG----	301
fscA	LTAGAALPPALV--RDVREALDRTL	VVWGMSEAGNGTSS-LS---	AD-APEVVSRSVGR	331
pcbAB	ILVGENLTEARYLA--LRQRFKNRILNEYGFTE	SAFVTALK-----	IFDPESTRKDTSLGR	319
→ bpsA	FSGGEALSRLAIQ-TTQEMPGRALIN	VGPTETTINS-SSFPVDPADLDEGPQ	SISIGS	318
fkbP	CQGGEALALDARLREL	CRHRPLRVHNHYGPAESQLITGYTL---	PADPDTWPAAPIGR	311
	*	:	.	:* :*

Figure 5.12. Sequence alignment of BpsA indicating the conserved glutamine involved in the ATP binding motif. The A domain of BpsA was identified by using the prediction tool on AntiSMASH. Following with that, other A domains from other biosynthetic gene clusters were identified by using MiBiG library across a range of different organisms, including: DifD from *Bacillus velezensis*, strain KKLW (GenBank: CP054714.1); HasG from *Anabaena* sp. 90 chromosome chANA01 (GenBank: CP003284.1); FscA from *Streptomyces* sp. FR-008 (GenBank: AY310323.2); PcbAB from *Penicillium chrysogenum* strain Wisconsin 54-1255 (GenBank: AM920436.1); FkbP from *Streptomyces* sp. MA6548 (GenBank: AF082100.1). Sequence alignment of these chosen A domains and the A domain of BpsA was determined by using the multiple sequence alignment tool from Clustal Omega.

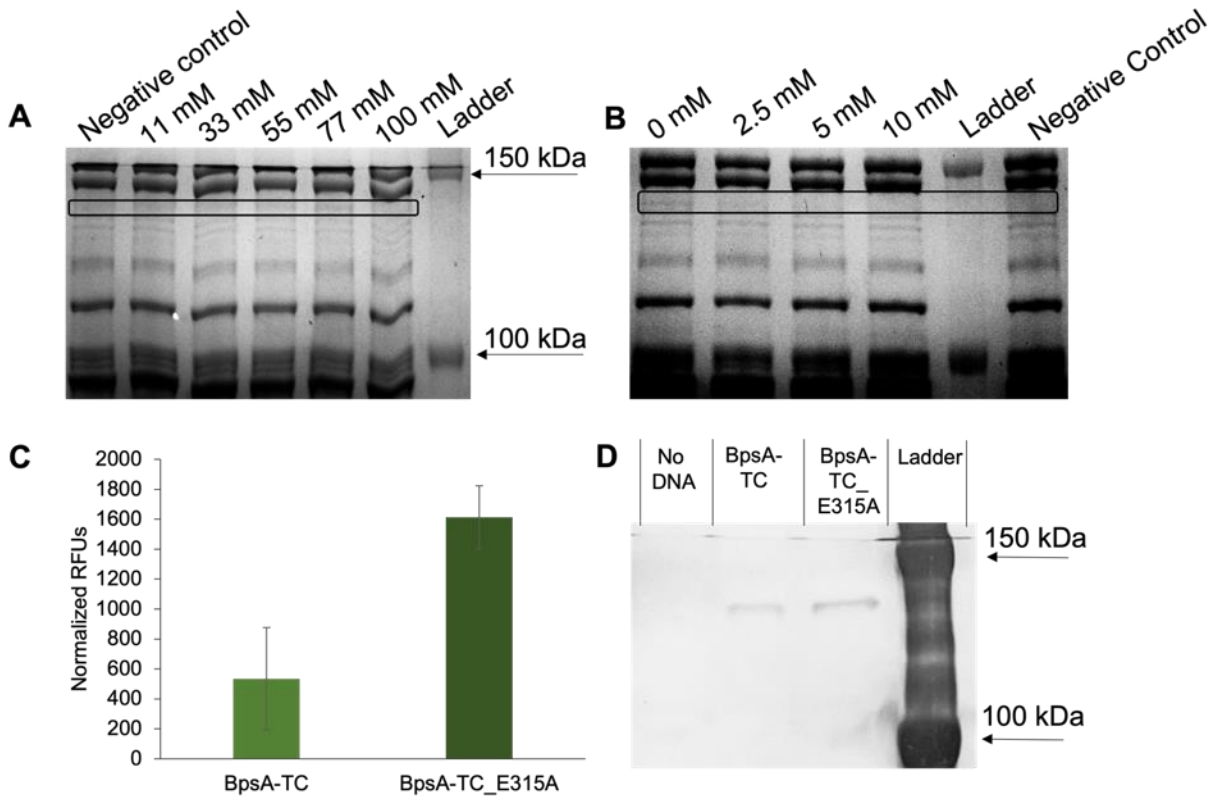


Figure 5.13. Effects of A domain activity on BpsA's PEP usage and expression. SDS-PAGE gels of pooled reactions (n=3) expressing BpsA-TC prepared with different concentrations of **A)** PEP and **B)** ATP. **C)** Differences in TC/FIAsH signals of BpsA-TC and BpsA-TC_E315A expression in reactions initiated with 77 mM PEP. **D)** Western Blot confirmation of BpsA-TC and BpsA-TC_E315A made in CFE reactions initiated with 77 mM PEP. All reactions were prepared with *E. coli* BL21 (Star)DE3 lysates. Reactions were run in triplicate and TC/FIAsH measurements were obtained by adding 30 μ M FIAsH reagent to each reaction after ~ 20 h incubation. TC/FIAsH measurements are presented as % relative intensities for visual comparison of trends between BpsA-TC and BpsA-TC_E315A data since these experiments were run on separate days. Error bars represent the standard error of the mean (n=3).

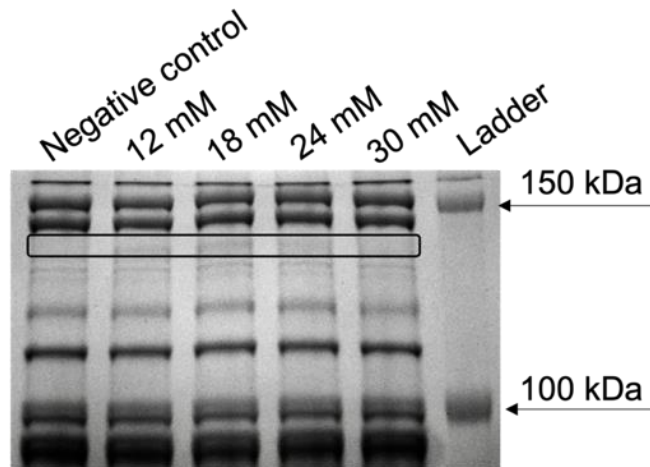


Figure 5.14. SDS-PAGE validation of magnesium concentration effects on BpsA expression. SDS-PAGE gel of pooled reactions (n=3) prepared with varying magnesium glutamate concentrations, expressing BpsA-TC. All reactions were prepared with *E. coli* BL21 (Star)DE3 lysates.

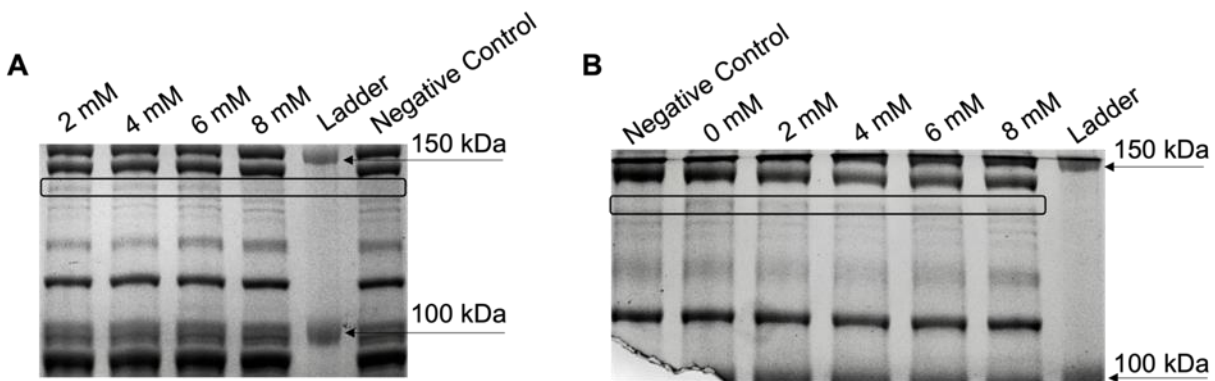


Figure 5.15. SDS-PAGE validation of amino acid concentration effects on BpsA expression. SDS-PAGE gels of pooled CFE reactions (n=3) expressing BpsA-TC with **A)** varying amino acids concentrations, and **B)** varying concentrations of supplemented L-glutamine. All reactions were prepared with *E. coli* BL21 (Star)DE3 lysates.

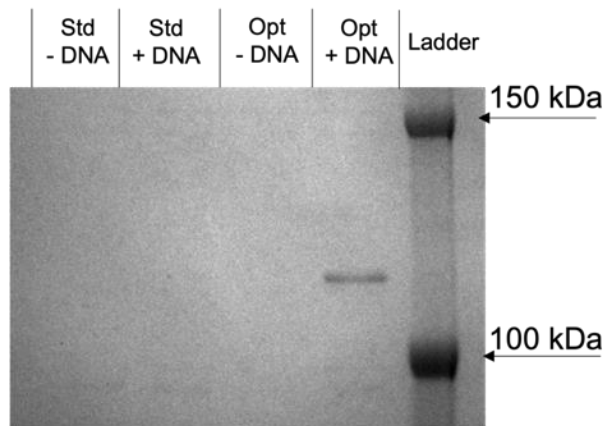


Figure 5.16. SDS-PAGE gel of elutions from the StrepII-tag purification of TycA. “Opt” = reactions run in BpsA optimized conditions (25 °C, 77 mM PEP, and 18 mM magnesium); “Std” = reactions run in standard conditions (30 °C, 33 mM PEP, and 12 mM magnesium).

Table 5.1. Plasmid and gene sequences used in this study.

Name	Sequence
<i>bpsA E. coli</i> codon optimized part 1	atgacgttgcaagagacttcggtactggagcctacctaaggggacaacgacctaaccggcttacttgcacagcgtg tcgctgaacaccagaagctattgcggtggcatactcgtgacgataagttaacctccgcgagttagctcccgcagcgc ggcgttggcagattacttagagcatttaggggttagcgcggatgactgttaggcctgtttagagccatccatcgacct gatggtgggcgcctggggaattttgaatgcaggagcagcttacttaccctttcgcggagtagcccgaggatcgtcttc gttatatgatcgagaattcggaaacaaaaattatttggcccagcaacgctcgtttagtcgctgcgcgagcttgcaccta aggacgttactattgtcacattgcgtgaaagcgaagcatttggccttaggggtactgaggcaccagctgccgtag cgcacgtccagacacccttgctatgtaactatactctggaatctacgggaaagccgaaggggtgatgattgagcatc gctccattgtaaacagtttaggatggctgcgcgaaacatatgccatcgatcgttcaaggctatcctcagaaaacccc gatgtctttgacgcagcgaatgggagatcctgtcgcgggtaacggtgctaccgtcgtaatgggcgccccaggagt tatgcagatccggaagggctgattgaaacaatcgtcaagcacaacgtcacaacattacagtggtcccaactctgtcgc agggttaattgatacagagaaattcccggagtgcgtatcactgcagcaaatttctcgggtggagaggcctaagtgc ctgtagcaattcaacaacccaggaaatgccgggacgtgcactgattaatgtatatggccccacggagacaacgat caatagtctagcttccagtcgatcccgcggatctggatgagggaccccaatcgatctcaatcggatcgctgtgcatg ggacgacttaccatatcttgacaagagactcttaagcctgtgggagtaggcgaaatcggggagttatacttgag gcattcagctggcccgcggctaccttcacgtgacgacttaactgccgaacgctttctgagattgaactggaggaagg gccgagccggtacgtttatacaagactggagacttaggccaatggaataatgacgggacggtccagttcgcgggtcgt gccgacaatcaagtgaacttcgtggatcgtgtggaattggacgagatctcttagctattgaaaatcacgattgggtg cgtaatgcggctgtaactcgtgaagaatgatgggcgacaggtttcagaatcttatcgctgcattgaattatccgagaa ggaagccgccttatggaccagggcaaccacggctcccaccatgcacccaagaaatcgaagtacaagtaaggct caactgtctaaccaggaactgcgtgacgacgcagaactggcagctcgtctgtttgacctggagggcgctgagcct acgccagaacagcgcgcccgtgtatttcccgcgaagactaccgctttacgaagggggagccgttacgcag gcggattacttgggcttctgggggtaccgttacggcaggggtattcgcgtaaaagccgcagatttggcgccagccgaatt aggacaaattctgcgttgggttggcagtagatctcgggaagaacgctgttgccaaagtatgggtatgcacccggggg ctttgtatgtacacagatgtactttgaacttgaagggtgtgggtgggctaaaccaggctactattactatcaaccgggtccgt catcaactgggtgcttattcagagcgtgaggctactgttaaggctaccgctcagattcactttattggcaaaaaatcgggc atcgagcctgtgtacaagaacaacatcctggaggttttagagatcgaaacgggtcacatggtaggacttttgaacaaa tttgcggcctacggattggacatccacgatcgtcttacaaccagctgtgaaggatttgcgtgatgagtgatgaag actattatcttgggacattcgagttgtgccacatgccggggcccgcatgaccaagccgaagtgtatgtccaaacgca tgggtgtaaaagtggccgactgccgaggggcaatatcgttacgaaaatggagaacttactcgcttttctgatgacattg tcttaagaagcatgttattgctatcaaccagtcctgtaccaggctgcatcg
<i>bpsA E. coli</i> codon optimized part 2	tttgaatttcagtatatagtcgcgcagaggaggaatggctgaaatatattacactgggcaaaaaactcaacatcttatg atgaacggctgaacttgggattcatgtccagcgggtattcgtccaagactggcaacccctgccagcttctgcgcgtatg gacgtgtattagcgcgaacggggtgactcggcccaatgtattttttagtaggagcgtatctctgatgagcagatt ggacatgaaggatgcgtgaagacagcgttcacatgcgcgggctgccgaattaatccgtgacgacttggctccttct gccagattatgatcccgaaccgtgtcgttgtgtcaccgtttaccgttagtgtaacggaaaaattgacgtcaaagc cttagctgcacagaccaagtcaacgcagaattgggtggagcgcggctcgtagcgcctcgtaccgagacagagaaag aaattgccgctgatgggagaaggctcttcgcgtgaaaatgcctcgggtcaagacgacttctcgaagtgaggtaa ttcgtgtattgcagtaggattagtcgcgaactgaatgcacgtttaggcgttagccttccctacagctgtactgagagc cctaccattgagaagtggctcgtcgttgaacgcgaagtcgcgcaggaatcgtcccgtttgtgcgttctcacgctga aaccggcaaaagcccgtccggtcatctgctggcctggctcgggaggtatccaatgaatttgcgctcacttgcaggagag atcggcctgggacgctcgttctatgggttcaaagctatggcatcaatgagggggagactccctacgaaaccattactg agatggcgaagaaagacatcgaggccttgaaggagatccagcccgcagggccataactctgtgggggtattctttcg gcgctcgtgtggcgtttgagactgcatacagctggaacaggcgggcgagaaagtggacaacttattctaatcggccc gggtagtccaaaagtccgtgcggagaatggcaagctcggggccgcgagggccagtttcccaatcggggtataca acgatccttttccgtcttactgggaccatcagcggcccagatcttgatcgtcgttagagacggtaactgacgaggca tccttcgccgaattcattagtgaactgaaaggcattgacgtggacttggcgccgctattattagtgctgtggccaacat atgaattcgaatactcgttccatgattggcgaacgtactttacaggcacctattagtttcaaggctgtgggtgacga ctattcgttccgtgaaaattcgagcggctatagcgtgagccaccgaccgtaattgacttggacgccgaccactatagctt gttgcgtgaagacatcggaagaactggtaagcatatccgctatttacttgggtgaatgttgcgggggtgtgtcatcaccat caccatcactaa

Table 5.1 continued

pys <i>E. coli</i> codon optimized part 1	atgcgcgatgtccacccatgccactgagtagtgcgcaacaagaagtgtgcttgccattagtcaatctcattcaaactc gaattaccatctttgtgatgtaattgaattgcgcggcgaacttcacttgccctggctgaagcggccattcggtccagtttt acaaacggacacccttcgcgcaacattcgacatcgacccgactacgggaacatacggcagcatatccagcctgcc gatgcgttacggcggaaggcatttgagtcacttgacgtaagcgaccaggcagaccccgcgaggccagtaacgagc tttgaggcatctgttgatcaagacatggattacaacacggggcgtaatccgctatgtgtgattcgtctggcaccgca acattatcgcatgattgaactggcttctcatttagtagtcgacggcttcggccacggaatccttttgggaatattactgtca ctataacgctttgtcacgcggtgagacggtagaagcactgaattagccccgttatctagcgtgttcgatgccaagaag agtatcgccattcgatgatgataagaaggatcgtagctattggcgccagattgtctgaaaatgccgaaccaacaca attggctccctggagatgtccgttgattaaactgaatcgtctgcgcaaagtctttggaggcgccacatttctcagctcgtg ccgcggtcgccgagcatcaactcgtttatctctatctgttggttattgtccacatattgcagcgcatgacggggca gcacgagttggccttgggcatgccagttgcggcacgtcagtaaggtcttgccaatgtacacctctatggtagcaaata tcttgccactgcacttgatcttacaccggagtcacacgtcttaagcgtcgcggccaatcttaacgccagttacgccatc actgtgtcatcaatcttaccgttcggagtcgatgatccgacgactacatgcggaacgcgggaataagccgttatttaata cctgttgaaacatcgttgatgaccaaggcgccgggatttgcgggctgcgatactacgatccaaaacgtcgccaacgg gccagccgaccaccttggtgacatcttcgacccgtcatgacgacggctgttgagatcggttcaatgccaatgctg atctgtattcggctgaggcattggagctgcactaccaacggttgactgcgttattcgagcgttcgtaacgctcccaaaa cgctggcagcagactataacgttcttaccagacgagcaacaacgccactacgcacttccccccatccagaaaag cttctgttttgcggaggcgttgcgaacgcgggtcatgactacgctgagcgccagcacttcccaaggagccagac cctgagttaccgctgttggacgatgacgctacacgcctggcagcccacctgcgcgaacgtgggtgcgcgcgggag attgcgtcgtatcatgttcagtcgctccgtagtggtggcgtgagcgccgtagcttggtaagctgggagcgtgttacgt accagtgatccagatctgcccgaagcccgattgaacatacttctctgacgcccagccggcagtagtcatgttgga ccaggtcacagcttaagggtgaggtagccgcagataaactgctgcgtctttccgcagagacttagctcagttgccag ctgtcacgcaacctctgtcagcattcgatgctggcttgctgctatcttattatactagcggatctacagggaagcctaa ggcgctggagggtgacgcagcgcaactta
pys <i>E. coli</i> codon optimized part 2	gtgccatcgctcgtactgctatcaatgccgcccagctgcaaccaggcgcgcggtgtgcttcagttcatcgacgtggatt gacatgtctgttttagagattatgatgacgcttttagcgggggcccgaattagtaattacagacaaagtatcgagtgctcca gggaaagcgcttgctacactggttaagcgtgagggcatcaatttctgttaatgaccccgctcctgttgccgtgtcaccaa actgaggactttcccaggatacggcttaatgcttgaggggagcctgttacgcccgcctgtggcgcttttgccca ctgtcgtttgtgaacgtgtatggaccaacagaaaccagtttcgaacaagtaataacgcctgtatggagcgggggact tatctatcgggtccggccacggctaacactcgttgtatgtagtcgattgggcaacagcgttacttctccgggagcctgg ggggattgtttatcgggggtccgggtgtagcccggtgtaccgtaaccgcccgcacttaactgtaaggggttttagca gattgttagactctgcatcaacgatgtaccgtgcaggggacgggtgttcttcgacctgacggacgcattcattacctgg gtcgtcaagataatcaaattaagttacgcgggttgcgcatcgaattagacgagattaagaatgtctttaggatgtgcag gcgtagatgacgccacagttattttacgtgaattggggcacggaccagcgattgtgggttacgtggcgtctaaggattcc agtcttgatggtcaacatcttaagcaggcgctggggcgtcattaccccaacatatggtgccgtccgtaatatgtgtgta gcgagttcccttgaccccaatggttaaattagctgttgatcgtttgccagctccggcactgtatgatccactgaactgg ctccagctcaaaacgctgaggaggcgccatgtgcaagttgttgcggaggccttagattgcaacgaagtttgcgcaa ccagaattttttgaccttgccggccatagtttactgggttgacgttgtgagtcgtattaaggagcagttcgggattgcac ttggaattgccgattttctgtgcgccaacaccgcgccaattagcgcagcgcttacagacaaactcaggcgactctga cccgtttgatgcagtcttaaccctgcgcagtgaaaggttctcgcccacccctgtttgcatccatccaggggggggattgc gtggccctatgccggattattaccattcttgccggaagaccaaccaatgtacgcccttaacgcggattctcgctgatcc gactcgtgttatcgggagcttagatgaactggccgagatattacaacgcattgtagactgtcatcctgaaggctccta ccaactgcaggctgtcagtgagggaaccttgcttacgtatcggtgtatgcttaagcgcggtggcgtgagggtat ctttttatgcatgtttgactcatatccattgcagggaggcccgcttcgttgaaactgtatgatgcgatgattatctcacgc gacccgtgctattgtcgaactcctcgtgcgggattaaagggcctgaagtcgcaatggaagaagcttgggaagccg taaatgtgcgacgaattcttgacacgcctgttagacgattcaaaattgatgttagagtttagggcgacgcagcagatga ggatttaattggtgagctgtgttcttcgcgcgacaacagacattctcgccaagatgagcagcagccaggtctgtggg caccgtacatttctggtagttaatccagcacgacgtcgaggctcctcacgaatgctgttcaacgcagctactaga caatttgggaaggcgttcgttagaggcttacttaacgctcaagcctccccgcctcggaaccccaaaagctgtcggaag tctgttgcgggggtgtgttgaggccaccgcagttcgaaaaataa

Table 5.1 continued

<i>tycA</i> <i>E. coli</i> codon optimized part 1	atgcttgcaaaccaagctaattctattgacctgaatgggagaaagaacaagttgcgcttttaacgacaccaaaagcgg agtaccctcaggacaagagcatccaccaattattcgaggaacgtgccgaagcagtcgccatcgctggcaattgtgt atgaaaatgagcgcttaacctaccaggaattaaataaaaaagccaaccaattagcgctgccctgattgaaaaaggg atcacacagattctatcgtagtgaggatgatggaagtgctattgaaaatgtcacagccattttagctattttaaggca ggtggagcgtagtctccatcgacattgagtagccgcgtgaccgtatccaatacatctgcaggactcccagactaaaat cgtaattacacaaaagaatgttagtcaattagctacgatgtcgggtatcgtagggcgtgtgtcgattggaagaagaaca gctggactctcgtcctaactcgaacttgcaattgcccgtcgaaccgactgatatggcctacgtaattataccctcaggcac aacaggcaagccaaagggaactatgttgaacacaaagggtattgcaaactgcagcttttttcagaacacggttgac gtaacgcccgcgaccgtattggtcagttcgcgtccatgagtttcgatgcaagtgtctgggagatgttatggctctttgac cggagccagcttatataattttgctgaacagacgattcatgactcatcagcttgaaaactaccttaatgagaatgagctt accattattactcttccccgacataccttacacacctaattcctgacctattactactcttcgatcatgattacggccgga tctgcagcaaaatttccgtagtaaatcggttggaataaaagtcggttatgtaaacgcctatgggcctacggaaacaag catttgtgcgactatttgggaggcgcccttaacgaacttttaaatcagacaatccctatcggaacccattcagaatac atatgcataattgttaacgaagaactgcaacttcagccattggctcggagggaattgtgcacgggggagtaggg ctggcccgcgggtattggaaccgtccgatttaaccgcggagaaagttgtgataatccttcgtccaggggaaaaaat gtatcgtaggggacgtggccaaatggctccgtcggagagatcgaattcctggccgtatcgaccaccagggtcaag atccgtgggaccgcacgtgagttggcgaaattgagctgtattgcttaacatgaacaaattaagggaagctgtcgtaat tgcgcggaagaccaacatgcgcaaccatattgtgcatacttcatttcgcagaaagaagtgcgcccgtcgaatt cgtgaatacgcgcgtcagaagctgccagcatacatgttgccgtcttatttcgttaacttgacaagatgccgtgacacct aatgataagattgaccgcaaggccctg
<i>tycA</i> <i>E. coli</i> codon optimized part 2	cccgaacctgacgtaacgtcggaactaatccagcgtatgagcctcctcgcaatcagaccgaaaagattctggctac agtatggcaggaagtcttggaaatcgaagattgggattctggacaactctattcgttaggtggtgatagcattcaagc gattcaagtggcagcgctctgcacgcctatcagcgcaagttagacacgaaggatttataactatccgaccattccc aagtcgcattatacgtgaagtcacgacccgcagctccgaacagggaattatcgaaggcccagtagcactgacacc catccagcactgggttttcgagaaaaacttcacaaacatgtcgcattggaaccaaagctatgttctgtatcgtgaacagg ggttcgatccagaggccattcgcaaagctctggataagattttaagtcaccacgacgcgttgccatgatctaccagcgt gaaaatgggtctatcgtacagcagaatcgcggttggttagtgaactgtatcaattctgcaatatgacctgacatcgca ctgggacgtacagaaggctatcgaagaagaaacaaacgcttgcatagttcaatgaatcgaagaagggaatta gtaaaagcgccctttccacacgcttaaggcgatcactgttcttgaatccaccatctggttatggacggtatttcttg gcgcattctgttgaagaccttgcaacggcatcacgtcaggtagttgcggggaaggagatcgttctgcccagaaaac tgattcatttaaggactggacagtggtgtaaacgattacgtaataagcagcgtctgatgcgtgaaatcccctactggga gaacttggaatcaggagctcgaatgcctcactccaaagactatgagacagtaggatgtaagcaaaaaagtacc gcaatatccaggttggtcttctgagcgtgaaaccgaacaattactgaaacaggcaaatagcgttaccaaacggaa attaatgatctgtttagccgcctgggtcttgagtggtgagtggtgggacgtctgacaaattgttatcaacttggaag ggcacggtcgtgaagacgttattgacacgcaaatgtgactcgtactattggctggtttacatctcagtatccggtccttct ggacgtctctcaaatgaaccacttaccaaaccacattaagtgtacaaaggagaatcttcgcaaaatcctaacaagg gtatcggctacgaaatctaaagtatatgacggaacccgaacgtcgcggtctttgtcattctcttgcacacctgaggttac attcaattatttaggtcagtttgattcaggggtgaattcagaactgttactcgtagccatactccagcggaatacccttg agccgatggtaagaataatcttagtcggattctgagattacactgccctgaacattaccggacgcattgaagggtggag agcttttgatcacatttactgattccgaggagcagtttcgtgaggagtcacatcaattatcggaataatcagaaac acctgcgtcgtacattacacactgcgtacaaaaaaagaagtagagcgtacaccctcggtttctccctgaagggcc tcagatggagggaatggatgatctttgaggtcttggttaacaccttaactgttgcggggtgtgttgagccacc gcagttcgaaaaataa

Table 5.1 continued

pET28a(+) sequence with TC tag	gatgtaactgaatgaaatggggaaggacgggtccagtaggctgcttcggcagcctactgttgagtagagtgtagctc cgtaactagttacatcctagcataacccttggggcctctaaccgggtcttgagggggtttttgcggaatccgaactcgagc accaccaccaccactgagatccggctgtaacaaagccgaaaggaagctgagttggctgtgccaccgctga gcaataactagcataacccttggggcctctaaccgggtcttgagggggtttttgctgaaaggaggaactatccggatt ggcgaatgggacgcccctgtagcggcgcaatgaagcggcggggttggtgttacgcgcagcgtgaccgtacact tgccagcggcctagcggcgtccttctgcttctcccttcttctcgccacgttcggcggttccccgtaagctctaaat cgggggctcccttaggggttcgatttagtgccttacggcacctcgaccccaaaaaactgattagggtgatggttcacgt agtgggccatcgccctgatagacgggttttcggccttgacgttgagtcacgttcttaatagtgactctgttccaaact ggaacaacactcaaccctatctcggctattctttgattataagggtatttgcgatttcggcctattggttaaaaaatgag ctgatttaacaaaaatgaacgcgaatttaacaaaatattaacgtttacaatttcagggtggcacttttcggggaaatgtgcg cgggaaccctatttgtttattttctaataacattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcat caaatgaaactgcaattattcatatcaggattatcaataaccatattttgaaaaagccggttctgtaataaggagaaaaac tcaccgaggcagttccataggtaggcaagatcctggatcggtctgcgattccgactcgccaacatcaataaacctat taatttcccctcgtaaaaaataagggtatcaagtgagaaatcacctagagtgacgactgaatccgggtgagaatggcaa aagttatgcatcttccagactgttcaacaggccagccattacgctcgtcatcaaaatcactcgcataaccaaacc gttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaggacaattacaaacaggaatcga atgcaaccggcgaggaacactgccagcgcatcaacaatatttcacctgaatcaggatattcttaataacctggaatg ctgtttccggggatcgcatggtgagtaacatgcatacagagtagcgaataaagtctgatggtcgggaagagg cataaattccgtcagccagtttagtctgaccatcctatctgaacatcattggcaacgctacctttgcatgttcagaaaca actctggcgcatcgggctccatacaatcgatagattgtcgacctgattgccgacattatcgcgagccattataacc cataaatacagcatccatgttgaatttaatcgcgccctagagcaagacgtttccggtgaatatggtcataacacccc ttgtattactgtttatgaagcagacaggttttattgtcatgacaaaatccctaacgtgagtttcgttccactgagcgtcaga ccccgtagaaaagatcaaaaggatcttctgagatcctttttctgcgcgtaactctgctgccaacaaaaaaaccacc gctaccagcggtggttgtttgcccgatcaagagctaccaactcttttccgaaggtaactggctcagcagagcgcgaga taccaaatactgtccttctagtgtagccgtagttaggccaccactcaagaactctgtagaccgcctacatacctcgctct gctaactctgttaccagtggctgctgccagtgggcgataagtcgtgtcttaccgggttgactcaagacgatagttaccgg ataaggcgagcggtcgggtgaacgggggttcgtgcacacagcccagcttgagcgaacgacctacaccgaac tgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaaggagaaaggcggaacaggtatccggtaa gcggcaggggtcggaaacaggagagcgacagggagctccaggggaaacgctggtatctttatagtcctgtcgg gtttcgccacctgactgagcgtcgattttgtgatgctgcagggggcgagcctatggaaaaacgccagcaac gcggccttttacgggtcctggcctttgtggtcctttgctcacatgttcttctcggttatccctgattctgttgataaccgta ttaccgctttgagtgagctgataccgctcgccgagccgaacgagcgcagcagcagtcagtgagcaggaagc ggaagagcgctgatgcggtattttccttacgcatctgtgcggtatttcacaccgcatatatggtcactctcagtacaat ctgctctgatgcgcgatagtaagccagatatacactccgctatcgtctacgtgactgggtcatggtgcgccccgacacc gccaacaccgctgacgcgcccgtacgggcttctgtctcggcgatccgcttacagacaagctgtgaccgtctccgg gagctgcatgtgtcagaggtttaccgctacaccgaaacgcgcgaggcagctgcggttaaagctcatcagcgtgggtc gtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttccagaagcgtaatgtctggctctgat aaagcggggccatgttaaggcggtttttctgtttggtcactgatgcctccgtgaagggggatttctgtcatggggtaa tgataccgatgaaacgagagaggtgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgtg agggtaaacaactggcggatggtgagggcgaccagagaaaaatcactcaggttaatgccagcgttctgtaat acagatgtagggttccacagggtagccagcagcatctgcgatgcagatccggaacataatggtgcagggcgctga ctccgctttccagactttacgaaacacggaacccaagaccattcatgttgtgtcaggtcgcagacgttttcagca gcagtcgttcacgttcgctcgcgtatcgggtattcattctgtaaccagtaaggcaaccccgccagcctagccgggttc tcaacgacaggagcagcatcatgcgcacccgtggggcgccatgcggcgataatggcctgctctcgccgaaacgt ttggtggcgggaccagtgacgaaggcttgagcagggcggtgcaagattccgaataaccgcaagcgacagggcgatc atcgtcgcgtccagcgaaagcggtcctcgccgaaatgaccagagcgtgcggcagcctgtcctacagagtgatg gataaagaagacagtcataagtgccgagacgatagtcacgcccgcgccaccggaaggtgactggtggtgaa ggctctcaagggtcagtcgagatcccggtgcctaagtagtgagtaactacattaattgcgtgtcgtcactgccc ctttccagtcgggaaacctgtcgtgccagctgctaataatgaatcgcccaacgcgcggggagaggcggttgcgtattgg gcgccaggggtgtttttctttaccagtgagacgggcaacagctgattgcccttaccgcctggcctgagagagttgc agcaagcggtccacgctgtttgtcccagcagggcgaaatcctgtttgatggtggttaacggcgggatataacatgag ctgtctcggtatcgtctatccactaccgagataccgcaccaacgcgcagccggactcggtaatggcgcgcatgtg cgccagcgccatctgatcgttggaaccagcatcgagtggaacgatgcctcattcagcatttgcatggtttgtga aaaccggacatggcactccagtcgcttccgctatcggtgaatttgattgcgagtgagatatttatgccagcca
--------------------------------------	---

Table 5.1 continued

pET28a(+) sequence with TC tag (continued)	gccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtagcccaatgcgacc agatgctccacgccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgctggtcagagacatcaa gaaataacgccggaacattagtgaggcagcttcacagcaatggcatcctggatccagcggatagttaatgatca gcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacac caccacgctggcaccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgctgcagggccaga ctggaggtggcaacgccaatcagcaacgactgtttgccgccagttgttgccacgcggttgggaatgtaattcagctc cgccatcgccgcttccacttttcccgcttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggctgata agagacaccggcatactctgcgacatcgataacgttactggtttcacattcaccaccctgaattgactcttccgggcg ctatcatgccataccgcgaaaggtttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcatt aggaagcagcccagtagtaggttaggccgttgagcaccgccgccaagggaatggtgcatgcaaggagatggcg ccaacagtcccccgccacggggcctgccaccataccacgccgaaacaagcgctcatgagcccgaagtggcg agcccgatcttcccatcgggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggcca cgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactatagggggaattgtgagcgga taacaattcccctctagaaataattttgtttaactttaagaaggagataatacc
---	--

Table 5.2. Primers used in this study; highlighted nucleotides are for the point mutation in the A domain: E315A.

Primer name	Sequence
<i>E. coli</i> BpsA part 1 Forward	CACACCAGGTCTCACAATGACGTTGCAAGAGACTTCGG
<i>E. coli</i> BpsA part 1 Reverse	CACACCAGGTCTCAAACGATGCAGCCTGGTACACGG
<i>E. coli</i> BpsA part 2 Forward	CACACCAGGTCTCAAACGATGCAGCCTGGTACACGG
<i>E. coli</i> BpsA part 2 Reverse	CACACCAGGTCTCATTATTCACCAAGTAAATAGCGGATATGC
pET28a with <i>E. coli</i> BpsA Forward	CACACCAGGTCTCAATAACTGCCAGCGCATCAACAATATTTTCACC
pET28a with <i>E. coli</i> BpsA Reverse	CACACCAGGTCTCAATAACTGCCAGCGCATCAACAATATTTTCACC
BpsA-TC_E315A part 1 Forward	AGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGT
BpsA-TC_E315A part 1 Reverse	ATCGTTGTGCGCGTGCGGCCATATACATTAATCAGTGCA
BpsA-TC_E315A part 2 Forward	GTATATGGCCCCACGGCGACAACGATCAATAGTTCTAGCTTTCCAGT
BpsA-TC_E315A part 2 Reverse	CGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTCTACGGG GTCTGACG
<i>E. coli</i> Pys part 1 Forward	CACACCAGGTCTCAGCATGCGCGATGTCCACCCATGC
<i>E. coli</i> Pys part 1 Reverse	CACACCAGGTCTCAACTAAGTTGCGCTGCGTCACCTCC
<i>E. coli</i> Pys part 2 Forward	CACACCAGGTCTCATAGTGCCCATCGCTCGTACTGC
<i>E. coli</i> Pys part 2 Reverse	CACACCAGGTCTCATTTTATTTTTCGAACTGCGGGTGGCTCC
pET28a with <i>E. coli</i> Pys Forward	CACACCAGGTCTCAAAAAGCTTGCGGCCGCACTCG
pET28a with <i>E. coli</i> Pys Reverse	CACACCAGGTCTCAAAAAGCTTGCGGCCGCACTCG
<i>E. coli</i> TycA part 1 Forward	CACACCAGGTCTCACAGCATGCTTGCAAACCAAGCTAATCTTATTGACCT
<i>E. coli</i> TycA part 1 Reverse	CACACCAGGTCTCAGGCAGGGCCTTGCGGTCAATCT
<i>E. coli</i> TycA part 2 Forward	CACACCAGGTCTCAGGCAGGGCCTTGCGGTCAATCT
<i>E. coli</i> TycA part 2 Reverse	CACACCAGGTCTCATTTTATTTTTCGAACTGCGGGTGGC
pET28a with <i>E. coli</i> TycA Forward	CACACCAGGTCTCAAAAAGCTTGCGGCCGCACTCG
pET28a with <i>E. coli</i> TycA Reverse	CACACCAGGTCTCAGCTGCTGCCCATGGTATATCTCC

CHAPTER 6

CONCLUDING STATEMENTS

Lysate-based cell-free systems offer a potential route to the accelerated growth of bio-based industries and, consequently, the bioeconomy. Economically viable cell factories are replacing a few petrochemical-based industries, but their optimization is a rate limiting process. Because cells are naturally resistant to being programmed, engineering their metabolic pathways to convert inexpensive substrates to value-added compounds is laborious and time-consuming. Integrating innovative strategies that improve the production efficiencies of lysates can help attain economically viable, next-generation manufacturing in these systems¹¹. Further, they afford easier and faster pathway building. This work provides techniques that can improve *E. coli* lysate pathway yield, a metric that must be optimized for sustainable production in cell extracts. These approaches are designed to be void of source strain viability constraints and to exploit features of cell-free systems that allow fast pathway engineering. Specifically, this work details the development of new strategies for 1) restructuring endogenous pathways post-lysis to focus flux through target native pathways, and 2) quickly improving the CFE of commercially relevant biosynthetic enzymes to increase flux through a heterologous module without source strain optimization. The results of these efforts support the feasibility of building efficient lysate-based biomanufacturing systems while providing points of improvement for achieving sustainable, totally cell-free engineering workflows.

Part I of this dissertation encompasses Chapters 2-4 and describes the development and application of a proteome engineering technique for maximizing conversion yield through pathways that are endogenous to the *E. coli* lysate. Part II of this thesis consists of Chapter 5 which implements a reporter-based assay for investigating biosynthetic enzyme expression under different CFE condition sets in lysates, enabling increased heterologous enzyme expression.

Chapter 2 details methods for analyzing lysate metabolomes and investigates how carbon flows from glucose through *E. coli* lysate metabolism. Given the less dynamic nature of lysate proteomes, metabolic flux in lysates is expected to behave differently compared to that in cells. The ability of interfacing pathways to “sense” one another via transcriptional regulation is lost in

lysates, resulting in potential bottlenecks or the unexpected accumulation of certain metabolites. Further, cofactors are present in limited supply which impacts the turnover of cofactor-regenerating reactions. By leveraging LC-MS/MS-enabled isotope tracing and HPLC techniques, metabolites formed in glucose-fed *E. coli* lysates can be measured through time. Based on these data, a model of the central metabolic network in this extract is proposed. The results confirm previous assumptions of metabolic behavior in lysates while contradicting others. For example, these results justify that lysate metabolism is largely fermentative, mostly producing lactate and ethanol, to regenerate cofactors for glycolysis. Succinate seems to be made largely through a fermentative pathway and other TCA cycle intermediates are undetected, implying an inactive TCA cycle in the *E. coli* extract. These data further provide important information on flux bottlenecks in glucose-fed extracts, given that several intermediates of central metabolic pathways are accumulated. Generally, the proposed model serves as a guide for future investigations of lysate metabolism. In the context of pathway engineering, insights obtained from these analyses provide design principles for restructuring the endogenous metabolism in lysates.

Chapter 3 introduces a new approach for the post-lysis reduction of selective enzyme activity from an *E. coli* extract. Pathway crosstalk inherent to cells still occurs in their derived lysates. Prior to the publication of this work, CFME efforts to shave the lysate proteome's complexity involved knocking out genes in the source strain^{47,53}. However, cell-free systems present the unique advantage of bypassing challenges in cellular engineering that are associated with host viability. Our method leverages the extract's open reaction environment for restructuring the lysate metabolic proteome after cell lysis. As a result, pyruvate can be transiently accumulated 40-fold better in a modified extract compared to the control without imposing metabolic burden on the host cell. This work provides a proof-of-concept that engineering the endogenous enzymes of cells to display tags can be innocuous to cell growth, and that the post-lysis manipulation of tagged enzymes is a viable cell-free engineering technique for reducing pathway crosstalk. Alternative approaches that are agnostic to source strain health involve the CRISPR-enabled down-

regulation of select genes or the use of inhibitors to inactivate certain enzymes in the lysate^{62,153}. In theory, gene down-regulation in the source strain cannot completely eliminate flux-competing reactions in derivative lysates. The use of inhibitors is restricted by their availability and cost of production. Meanwhile, the affinity-based pull-down procedure can be optimized to improve the removal of select enzymes. These results could also inspire the exploration of other protein tag technologies that would reduce the method's overall cost.

Chapter 4 leverages insights gained from Chapter 2 and the method established in Chapter 3 to develop the first workflow that engineers 100% yield through a lysate pathway that initially had 10% yield. This was achieved by the optimization of our proteome engineering method to remove unwanted reactions, and its application with complementary CFME techniques that drive flux through bottlenecks and towards a specific native end-product, ethanol, in *E. coli* lysates. The workflow is completely cell-free as every step is unconstrained by source strain viability, allowing pathway yield optimization while bypassing time-consuming source strain optimization. Importantly, demonstrating maximum conversion yield in lysates advances the feasibility of their commercialization for synthesizing high-value targets especially when coupled with efficient heterologous pathway modules. Furthermore, the technique can be used for prototyping high-yield or even unconventional pathway designs in integrated *in vitro-in vivo* engineering frameworks.

Future efforts will seek to address limitations of the flux rewiring workflow developed in Part I, which include the efficiency of the genome engineering tool used and the cost of the current enzyme depletion strategy. While pORTMAGE's portability and specificity make it an appealing choice for genome editing, its efficiency for inserting protein tag sequences into the genome is low, requiring 23 cycles (~ 2 weeks) of MAGE for the insertion of the 6xHis-tag sequence at both *pflB* and *ldhA* sites. The incorporation of genome engineering tools with higher success rates per cycle will speed up this part of the enzyme depletion process. Since we initiated the development of our strategy, pORTMAGE has been modified to allow 5- to 10-fold increases in multiplex editing

in *E. coli*²⁴⁷. A CRISPR-based technique for integrating up to 0.5 kb DNA into targeted sites with ~ 80% efficiency in *E. coli* has also been released²⁴⁸. Alternative protein tag technologies can be assessed to reduce the cost of the enzyme removal process which, in its current state, is too expensive for building profitable lysate-based biomanufacturing systems. We discuss the economics of this process and the feasibility of using cellulose-binding domains and microcrystalline cellulose beads in Chapter 4 (Appendix). The genomic integration of protease tags and compartmentalized expression of a compatible protease in the same source strain could also be cheaper than an affinity-based enzyme removal approach.

Part II of this dissertation, which consists of Chapter 5, details the optimization and implementation of a reporter-based assay for quickly sampling different condition sets in lysate-based CFE reactions. CFE is a powerful tool for constructing heterologous pathways at sufficient concentrations for high-yield bioconversion, especially when considering enzyme classes that are difficult to synthesize or functionalize *in vivo*. These include naturally abundant biosynthetic enzymes that convert central metabolites into high-value natural products²⁸. Broadly, this work investigates how target enzyme structure and catalytic activity impact CFE requirements in extracts. The resource competition between TX-TL processes involved in the synthesis of the enzyme and the enzyme itself is highlighted as a significant barrier in the advancement of CFE for rapid heterologous pathway building in lysates. Tailoring CFE reaction condition sets to different enzyme targets is thus needed, especially for enzymes that interface with lysate endogenous metabolism and thus contribute to the unproductive sequestration of TX-TL resources (e.g., energy substrates and salts). This tool enables such investigation, demonstrated through the improved expression of large biosynthetic enzymes, NRPSs, in cell extracts with higher throughput than conventional techniques. The results also show a concordant increase in nonribosomal peptide formation with higher NRPS expression. By accelerating CFE condition optimization for valuable biosynthetic enzymes, the current strategy enables the rapid construction of efficient whole biosynthetic pathways in extracts. This is not only important for

quickly expressing known heterologous enzymes, but also for screening engineered variants or novel heterologous pathways²⁴⁹.

Future CFME studies that leverage CFE for heterologous pathway building will benefit from refining the current technique or exploring alternatives for high-throughput enzyme expression validation. A limitation of the FIAsh reagent is that it interacts with the lysate background and possibly different components of the reaction environment. Consequently, the signal-to-noise ratio decreases over time, and only end-point measurements can be reliably made. Further, careful negative controls must be designed to subtract noise when measuring TC/FIAsh fluorescence under different reaction condition sets. Testing tags that carry the tetracysteine motif but have higher specificity for FIAsh may reduce this noise. An alternative approach that can be used for screening reaction conditions was reported at the same time we ended our investigation. This method leverages the split-GFP system and was used to validate the parallel expression of enzyme homologs in lysates²⁵⁰. Briefly, a 28-amino acid tag carrying a small, non-fluorescent fragment of sfGFP called GFP11 is displayed on the C-terminus of an expressed enzyme in a lysate reaction. A larger fragment of sfGFP called GFP1-10 is then added exogenously to the enzyme enriched lysate and forms the fluorescent GFP complex with the tagged enzyme. The split-GFP assay does not suffer from high background noise and thus does not require negative (i.e., no DNA) controls when making relative comparisons of different reaction condition sets. However, GFP11 is a larger tag than the 6-amino acid TC tag and may thus interfere with the structure and activity of the tagged enzyme. Also, whether different reaction conditions (e.g., ionic strength, reaction temperature) impact the binding of GFP1-10 and GFP11 has yet to be evaluated.

Developing strategies that improve the production efficiencies of lysates while affording accelerated (i.e., unconstrained by cell viability) engineering campaigns will help actualize lysate-based biomanufacturing. The expedited development of biomanufacturing processes in lysates can accelerate the diversification of bioindustries and the subsequent replacement of most

petrochemical-based manufacturing processes with sustainable alternatives. Additionally, the ability of productive extracts to be lyophilized and distributed without a cold chain enables decentralized manufacturing during supply chain disruptions. Thus, refining and integrating these strategies into cost-effective CFME workflows will expand our biomanufacturing capabilities, and as our cell-free engineering toolkit grows, so too will the bioeconomy.

REFERENCES

- (1) National Academies of Sciences, E. and M. (2020) Safeguarding the Bioeconomy. *Safeguarding the Bioeconomy*. National Academies Press.
- (2) Gallo, M. E. (2022) The Bioeconomy: A Primer.
- (3) Rogers, J. N., Stokes, B., Dunn, J., Cai, H., Wu, M., Haq, Z., and Baumes, H. (2017) An assessment of the potential products and economic and environmental impacts resulting from a billion ton bioeconomy. *Biofuels, Bioprod. Biorefining* 11, 110–128.
- (4) Hinman, A. W., and Friedman, D. (2022) Actions to Enable an Equitable and Innovative. *Ebrc*.
- (5) Chui, M., Evers, M., Manyika, J., Zheng, A., and Nisbet, T. (2020) The Bio Revolution Innovations transforming economies, societies, and our lives. *McKinsey Glob. Inst.*
- (6) Scown, C. D. (2022) Prospects for carbon-negative biomanufacturing. *Trends Biotechnol.* 40, 1415–1424.
- (7) Scown, C. D., and Keasling, J. D. (2022) Sustainable manufacturing with synthetic biology. *Nat. Biotechnol.* 2022 403 40, 304–307.
- (8) Tan, E. C. D., and Lamers, P. (2021) Circular Bioeconomy Concepts—A Perspective. *Front. Sustain.* 2, 53.
- (9) (2018) Fourth Industrial Revolution for the Earth Series Harnessing the Fourth Industrial Revolution for Life on Land.
- (10) Zhang, Y. H. P., Sun, J., and Ma, Y. (2017, May 1) Biomanufacturing: history and perspective. *J. Ind. Microbiol. Biotechnol.* Springer Verlag.
- (11) Rasor, B. J., Vögeli, B., Landwehr, G. M., Bogart, J. W., Karim, A. S., and Jewett, M. C. (2021) Toward sustainable, cell-free biomanufacturing. *Curr. Opin. Biotechnol.* 69, 136–144.
- (12) Du, J., Shao, Z., and Zhao, H. (2011) Engineering microbial factories for synthesis of value-added products. *J. Ind. Microbiol. Biotechnol.* 38, 873.

- (13) Woolston, B. M., Edgar, S., and Stephanopoulos, G. (2013) Metabolic Engineering: Past and Future. *Annu. Rev. Chem. Biomol. Eng.* 4, 259–288.
- (14) Zhang, J., Babbitt, A., and Stephanopoulos, G. (2012, November 1) Metabolic engineering: Enabling technology of a bio-based economy. *Curr. Opin. Chem. Eng.* Elsevier Ltd.
- (15) García-Granados, R., Lerma-Escalera, J. A., and Morones-Ramírez, J. R. (2019) Metabolic engineering and synthetic biology: Synergies, future, and challenges. *Front. Bioeng. Biotechnol.* 7, 36.
- (16) Julleson, D., David, F., Pfleger, B., and Nielsen, J. (2015) Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnol. Adv.* 33, 1395–1402.
- (17) Lim, H. J., and Kim, D. M. (2019) Cell-Free Metabolic Engineering: Recent Developments and Future Prospects. *Methods Protoc.* 2019, Vol. 2, Page 33 2, 33.
- (18) Guo, W., Sheng, J., and Feng, X. (2017, January 1) Mini-review: In vitro Metabolic Engineering for Biomanufacturing of High-value Products. *Comput. Struct. Biotechnol. J.* Elsevier B.V.
- (19) Swartz, J. R. (2018) Expanding biological applications using cell-free metabolic engineering: An overview. *Metab. Eng.*
- (20) Bowie, J. U., Sherkhanov, S., Korman, T. P., Valliere, M. A., Opgenorth, P. H., and Liu, H. (2020, January 23) Synthetic Biochemistry: The Bio-inspired Cell-Free Approach to Commodity Chemical Production. *Trends Biotechnol.* Elsevier Ltd.
- (21) Dudley, Q. M., Karim, A. S., and Jewett, M. C. (2015) Cell-Free Metabolic Engineering: Biomanufacturing beyond the cell. *Biotechnol. J.* 10, 69.
- (22) Karim, A. S., and Jewett, M. C. (2018) Cell-Free Synthetic Biology for Pathway Prototyping, in *Methods in Enzymology*, pp 31–57. Academic Press Inc.
- (23) Karim, A. S., Dudley, Q. M., Juminaga, A., Yuan, Y., Crowe, S. A., Heggstad, J. T., Garg, S., Abdalla, T., Grubbe, W. S., Rasor, B. J., Coar, D. N., Torculas, M., Krein, M., Liew, F. M. (Eric), Quattlebaum, A., Jensen, R. O., Stuart, J. A., Simpson, S. D., Köpke, M., and Jewett, M. C. (2020)

In vitro prototyping and rapid optimization of biosynthetic enzymes for cell design. *Nat. Chem. Biol.* 2020 168 16, 912–919.

(24) Dudley, Q. M., Karim, A. S., Nash, C. J., and Jewett, M. C. (2020) In vitro prototyping of limonene biosynthesis using cell-free protein synthesis. *Metab. Eng.* 61, 251–260.

(25) Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S., and Koffas, M. A. G. (2016) Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol.* 34, 652–664.

(26) Vögeli, B., Schulz, L., Garg, S., Tarasava, K., Clomburg, J. M., Lee, S. H., Gonnot, A., Mouilly, E. H., Kimmel, B. R., Tran, L., Zeleznik, H., Brown, S. D., Simpson, S. D., Mrksich, M., Karim, A. S., Gonzalez, R., Köpke, M., and Jewett, M. C. (2022) Cell-free prototyping enables implementation of optimized reverse β -oxidation pathways in heterotrophic and autotrophic bacteria. *Nat. Commun.* 2022 131 13, 1–10.

(27) 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.

(28) Bogart, J. W., Cabezas, M. D., Vögeli, B., Wong, D. A., Karim, A. S., and Jewett, M. C. (2021) Cell-Free Exploration of the Natural Product Chemical Space. *ChemBioChem* 22, 84–91.

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

(30) Silverman, A. D., Karim, A. S., and Jewett, M. C. (2020, March 1) Cell-free gene expression: an expanded repertoire of applications. *Nat. Rev. Genet.* Nature Research.

(31) Chong, S. (2014) Overview of cell-free protein synthesis: Historic landmarks, commercial systems, and expanding applications. *Curr. Protoc. Mol. Biol.* 2014, 16.30.1-16.30.11.

(32) Swartz, J. (2006) Developing cell-free biology for industrial applications. *J Ind Microbiol Biotechnol* 33, 476–485.

(33) Moore, S. J., MacDonald, J. T., and Freemont, P. S. (2017) Cell-free synthetic biology for in vitro prototype engineering. *Biochem. Soc. Trans.* 45, 785.

- (34) Rosano, G. L., and Ceccarelli, E. A. (2014) Recombinant protein expression in microbial systems. *Front. Microbiol.* Frontiers Research Foundation.
- (35) Gregorio, N. E., Levine, M. Z., and Oza, J. P. (2019) A User's Guide to Cell-Free Protein Synthesis. *Methods Protoc.* 2, 24.
- (36) 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.
- (37) Tsoi, R., Wu, F., Zhang, C., Bewick, S., Karig, D., and You, L. (2018) Metabolic division of labor in microbial systems. *Proc. Natl. Acad. Sci. U. S. A.* 115, 2526–2531.
- (38) Rosano, G. L., and Ceccarelli, E. A. (2014, April 17) Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front. Microbiol.* Frontiers Research Foundation.
- (39) Khambhati, K., Bhattacharjee, G., Gohil, N., Braddick, D., Kulkarni, V., and Singh, V. (2019) Exploring the Potential of Cell-Free Protein Synthesis for Extending the Abilities of Biological Systems. *Front. Bioeng. Biotechnol.* 7, 248.
- (40) 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.
- (41) Pardee, K. (2018) Perspective: Solidifying the impact of cell-free synthetic biology through lyophilization. *Biochem. Eng. J.* 138, 91–97.
- (42) Warfel, K. F., Williams, A., Wong, D. A., Sobol, S. E., Desai, P., Li, J., Chang, Y. F., DeLisa, M. P., Karim, A. S., and Jewett, M. C. (2023) A Low-Cost, Thermostable, Cell-Free Protein Synthesis Platform for On-Demand Production of Conjugate Vaccines. *ACS Synth. Biol.* 12, 95–107.
- (43) Wilding, K. M., Zhao, E. L., Earl, C. C., and Bundy, B. C. (2019) Thermostable lyoprotectant-enhanced cell-free protein synthesis for on-demand endotoxin-free therapeutic production. *N. Biotechnol.* 53, 73–80.
- (44) Bundy, B. C., Hunt, J. P., Jewett, M. C., Swartz, J. R., Wood, D. W., Frey, D. D., and Rao, G. (2018) Cell-free biomanufacturing. *Curr. Opin. Chem. Eng.* 22, 177–183.

- (45) Kay, J. E., and Jewett, M. C. (2020) A cell-free system for production of 2,3-butanediol is robust to growth-toxic compounds. *Metab. Eng. Commun.* 10, e00114.
- (46) 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.
- (47) Bujara, M., Schümperli, M., Billerbeck, S., Heinemann, M., and Panke, S. (2010) Exploiting cell-free systems: Implementation and debugging of a system of biotransformations. *Biotechnol. Bioeng.* 106, 376–389.
- (48) Cui, J., Stevenson, D., Korosh, T., Amador-Noguez, D., Olson, D. G., and Lynd, L. R. (2020) Developing a Cell-Free Extract Reaction (CFER) System in *Clostridium thermocellum* to Identify Metabolic Limitations to Ethanol Production. *Front. Energy Res.* 8, 72.
- (49) O’Kane, P. T., Dudley, Q. M., McMillan, A. K., Jewett, M. C., and Mrksich, M. (2019) High-throughput mapping of CoA metabolites by SAMDI-MS to optimize the cell-free biosynthesis of HMG-CoA. *Sci. Adv.* 5, eaaw9180.
- (50) Tang, S., Liao, D., Li, X., Lin, Y., Han, S., and Zheng, S. (2021) Cell-Free Biosynthesis System: Methodology and Perspective of in Vitro Efficient Platform for Pyruvate Biosynthesis and Transformation. *ACS Synth. Biol.* 10, 2417–2433.
- (51) Grubbe, W. S., Rasor, B. J., Krüger, A., Jewett, M. C., and Karim, A. S. (2020) Cell-free styrene biosynthesis at high titers. *Metab. Eng.* 61, 89–95.
- (52) Bujara, M., 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.
- (53) Dudley, Q. M., Anderson, K. C., and Jewett, M. C. (2016) Cell-Free Mixing of *Escherichia coli* Crude Extracts to Prototype and Rationally Engineer High-Titer Mevalonate Synthesis.
- (54) Zhuang, L., Huang, S., Liu, W. Q., Karim, A. S., Jewett, M. C., and Li, J. (2020) Total in vitro biosynthesis of the nonribosomal macrolactone peptide valinomycin. *Metab. Eng.* 60, 37–44.
- (55) Zawada, J. F., Yin, G., Steiner, A. R., Yang, J., Naresh, A., Roy, S. M., Gold, D. S., Heinsohn, H. G., and Murray, C. J. (2011) Microscale to manufacturing scale-up of cell-free cytokine

production-a new approach for shortening protein production development timelines. *Biotechnol. Bioeng.* 108, 1570–1578.

(56) Opgenorth, P. H., Korman, T. P., and Bowie, J. U. (2014) A synthetic biochemistry molecular purge valve module that maintains redox balance. *Nat. Commun.* 2014 51 5, 1–8.

(57) 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.

(58) King, E., Maxel, S., and Li, H. (2020) Engineering natural and noncanonical nicotinamide cofactor-dependent enzymes: design principles and technology development. *Curr. Opin. Biotechnol.* 66, 217–226.

(59) Black, W. B., and Li, H. (2022) Cell-Free Noncanonical Redox Cofactor Systems. *Methods Mol. Biol.* 2433, 185–198.

(60) Stephanopoulos, G., Aristidou, A., and Nielsen, J. (1998) Metabolic Engineering: Principles and Methodologies. Academic Press, San Diego.

(61) Khattak, W. A., Ul-Islam, M., Ullah, M. W., Yu, B., Khan, S., and Park, J. K. (2014) Yeast cell-free enzyme system for bio-ethanol production at elevated temperatures. *Process Biochem.* 49, 357–364.

(62) Rasor, B. J., Yi, X., Brown, H., Alper, H. S., and Jewett, M. C. (2021) An integrated in vivo/in vitro framework to enhance cell-free biosynthesis with metabolically rewired yeast extracts. *Nat. Commun.* 12.

(63) Jin, X., and Hong, S. H. (2018, October 15) Cell-free protein synthesis for producing ‘difficult-to-express’ proteins. *Biochem. Eng. J.* Elsevier B.V.

(64) Dopp, B. J. L., Tamiev, D. D., and Reuel, N. F. (2019) Cell-free supplement mixtures: Elucidating the history and biochemical utility of additives used to support in vitro protein synthesis in *E. coli* extract. *Biotechnol. Adv.* 37, 246–258.

(65) Koch, M., Faulon, J. L., and Borkowski, O. (2018) Models for cell-free synthetic biology: Make prototyping easier, better, and faster. *Front. Bioeng. Biotechnol.* 6, 182.

- (66) Wick, S., Walsh, D. I., Bobrow, J., Hamad-Schifferli, K., Kong, D. S., Thorsen, T., Mrozczyk, K., and Carr, P. A. (2019) PERSIA for Direct Fluorescence Measurements of Transcription, Translation, and Enzyme Activity in Cell-Free Systems. *ACS Synth. Biol.* 8, 1010–1025.
- (67) Kries, H. (2016) Biosynthetic engineering of nonribosomal peptide synthetases. *J. Pept. Sci.* 22, 564–570.
- (68) Ji, X., Liu, W. Q., and Li, J. (2022) Recent advances in applying cell-free systems for high-value and complex natural product biosynthesis. *Curr. Opin. Microbiol.* 67, 102142.
- (69) Brown, A. S., Robins, K. J., and Ackerley, D. F. (2017) A sensitive single-enzyme assay system using the non-ribosomal peptide synthetase BpsA for measurement of L-glutamine in biological samples. *Sci. Rep.* 7, 1–13.
- (70) 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.
- (71) Dudley, Q. M., Nash, C. J., and Jewett, M. C. (2019) Cell-free biosynthesis of limonene using enzyme-enriched Escherichia coli lysates. *Synth. Biol.* 4.
- (72) 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.
- (73) Mohr, B., Giannone, R. J., Hettich, R. L., and Doktycz, M. J. (2020) Targeted growth medium dropouts promote aromatic compound synthesis in crude E. Coli cell-free systems. *ACS Synth. Biol.* 9, 2986–2997.
- (74) Garcia, D. C., Dinglasan, J. L. N., Shrestha, H., Abraham, P. E., Hettich, R. L., and Doktycz, M. J. (2021) A lysate proteome engineering strategy for enhancing cell-free metabolite production. *Metab. Eng. Commun.* 12, e00162.

- (75) Coskun, O. (2016) Separation Techniques: CHROMATOGRAPHY. *North. Clin. Istanbul* 3, 156.
- (76) Bernardes, A. N., Godoi, L. A. G., Camiloti, P. R., Gomes, A. da C., Sanchez, B. L. S., Torres, A. P. R., and Botta, L. S. (2021) Organic acids and alcohols quantification by HPLC/RID in sugarcane vinasse: analytical method validation and matrix effect assessment. *Int. J. Environ. Anal. Chem.* 101, 325–336.
- (77) Garcia, D. C., Cheng, X., Land, M. L., Standaert, R. F., Morrell-Falvey, J. L., and Doktycz, M. J. (2019) Computationally Guided Discovery and Experimental Validation of Indole-3-acetic Acid Synthesis Pathways. *ACS Chem. Biol.* 14, 2867–2875.
- (78) Karim, A. S., Rasor, B. J., and Jewett, M. C. (2020) Enhancing control of cell-free metabolism through pH modulation. *Synth. Biol.* 5.
- (79) Xiao, J. F., Zhou, B., and Ressom, H. W. (2012, February) Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *TrAC - Trends Anal. Chem.* NIH Public Access.
- (80) Asensio-Ramos, M., Fanali, C., D’Orazio, G., and Fanali, S. (2017) Nano-liquid chromatography, in *Liquid Chromatography: Fundamentals and Instrumentation: Second Edition*, pp 637–695. Elsevier.
- (81) Nagana Gowda, G. A., and Djukovic, D. (2014) Overview of mass spectrometry-based metabolomics: Opportunities and challenges. *Methods Mol. Biol.* 1198, 3–12.
- (82) Jaishankar, J., and Srivastava, P. (2017, October 16) Molecular basis of stationary phase survival and applications. *Front. Microbiol.* Frontiers Media S.A.
- (83) Bartolomeo, M. P., and Maisano, F. (2006) Validation of a reversed-phase HPLC method for quantitative amino acid analysis. *J. Biomol. Tech.* 17, 131–137.
- (84) Hauck, T., Landmann, C., Brühlmann, F., and Schwab, W. (2003) Formation of 5-methyl-4-hydroxy-3[2H]-furanone in cytosolic extracts obtained from *Zygosaccharomyces rouxii*. *J. Agric. Food Chem.* 51, 1410–1414.

- (85) Huang, H., Yuan, M., Seitzer, P., Ludwigsen, S., and Asara, J. M. (2020) IsoSearch: An Untargeted and Unbiased Metabolite and Lipid Isotopomer Tracing Strategy from HR-LC-MS/MS Datasets. *Methods Protoc.* 3, 1–16.
- (86) Guterl, J. K., Garbe, D., Carsten, J., Steffler, F., Sommer, B., Reiß, S., Philipp, A., Haack, M., Rühmann, B., Koltermann, A., Ketting, U., Brück, T., and Sieber, V. (2012) Cell-free metabolic engineering: Production of chemicals by minimized reaction cascades. *ChemSusChem* 5, 2165–2172.
- (87) 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.
- (88) 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.
- (89) 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.
- (90) 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.
- (91) 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.
- (92) 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.
- (93) Krüger, A., Mueller, A. P., Rybnicky, G. A., Engle, N. L., Yang, Z. K., Tschaplinski, T. J., Simpson, S. D., Köpke, M., and Jewett, M. C. (2020) Development of a clostridia-based cell-free

- system for prototyping genetic parts and metabolic pathways. *Metab. Eng.* 62, 95–105.
- (94) 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.
- (95) Tinafar, A., Jaenes, K., and Pardee, K. (2019) Synthetic Biology Goes Cell-Free. *BMC Biol.* 17, 1–14.
- (96) 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.
- (97) Yin, G., Garces, E. D., Yang, J., Zhang, J., Tran, C., Steiner, A. R., Roos, C., Bajad, S., Hudak, S., Penta, K., Zawada, J., Pollitt, S., and Murray, C. J. (2012) Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system. *MAbs* 4, 217–225.
- (98) Oesterle, S., Roberts, T. M., Widmer, L. A., Mustafa, H., Panke, S., and Billerbeck, S. (2017) Sequence-based prediction of permissive stretches for internal protein tagging and knockdown. *BMC Biol.* 15, 1–17.
- (99) 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.
- (100) Maleki, N., and Eiteman, M. (2017) Recent Progress in the Microbial Production of Pyruvic Acid. *Fermentation* 3, 8.
- (101) Dai, Z., and Nielsen, J. (2015) Advancing metabolic engineering through systems biology of industrial microorganisms. *Curr. Opin. Biotechnol.* 36, 8–15.
- (102) 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.*
- (103) Loughheed, 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.

(104) 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.

(105) 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. U. S. A.* 113, 2502–7.

(106) 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.

(107) Soini, J., Ukkonen, K., and Neubauer, P. (2008) High cell density media for Escherichia coli are generally designed for aerobic cultivations - Consequences for large-scale bioprocesses and shake flask cultures. *Microb. Cell Fact.* 7, 26.

(108) Enfors, S. O., Jahic, M., Rozkov, A., Xu, B., Hecker, M., Jürgen, B., Krüger, E., Schweder, T., Hamer, G., O'Beirne, D., Noisommit-Rizzi, N., Reuss, M., Boone, L., Hewitt, C., McFarlane, C., Nienow, A., Kovacs, T., Trägårdh, C., Fuchs, L., Revstedt, J., Friberg, P. C., Hjertager, B., Blomsten, G., Skogman, H., Hjort, S., Hoeks, F., Lin, H. Y., Neubauer, P., Van der Lans, R., Luyben, K., Vrabel, P., and Manelius, Å. (2001) Physiological responses to mixing in large scale bioreactors. *J. Biotechnol.* 85, 175–185.

(109) Wang, Q., Ou, M. S., Kim, Y., Ingram, L. O., and Shanmugam, K. T. (2010) Metabolic flux control at the pyruvate node in an anaerobic escherichia coli strain with an active pyruvate dehydrogenase. *Appl. Environ. Microbiol.* 76, 2107–2114.

(110) Xu, B., Jahic, M., Blomsten, G., and Enfors, S. O. (1999) Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with Escherichia coli. *Appl. Microbiol. Biotechnol.* 51, 564–571.

- (111) Foshag, D., Henrich, E., Hiller, E., Schäfer, M., Kerger, C., Burger-Kentischer, A., Diaz-Moreno, I., García-Mauriño, S. M., Dötsch, V., Rupp, S., and Bernhard, F. (2018) The *E. coli* S30 lysate proteome: A prototype for cell-free protein production. *N. Biotechnol.* 40, 245–260.
- (112) Snoep, J. L., de Graef, M. R., Westphal, A. H., de Kok, A., Joost Teixeira de Mattos, M., and Neijssel, O. M. (1993) Differences in sensitivity to NADH of purified pyruvate dehydrogenase complexes of *Enterococcus faecalis*, *Lactococcus lactis*, *Azotobacter vinelandii* and *Escherichia coli*: Implications for their activity in vivo. *FEMS Microbiol. Lett.* 114, 279–283.
- (113) Murarka, A., Clomburg, J. M., Moran, S., Shanks, J. V., and Gonzalez, R. (2010) Metabolic analysis of wild-type *Escherichia coli* and a Pyruvate Dehydrogenase Complex (PDHC)-deficient derivative reveals the role of PDHC in the fermentative metabolism of glucose. *J. Biol. Chem.* 285, 31548–31558.
- (114) Wagner, A. F. V., Frey, M., Neugebauer, F. A., Schafer, W., and Knappe, J. (1992) The free radical in pyruvate formate-lyase is located on glycine-734. *Proc. Natl. Acad. Sci. U. S. A.* 89, 996–1000.
- (115) 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.
- (116) Ronda, C., Pedersen, L. E., Sommer, M. O. A., and Nielsen, A. T. (2016) CRMAGE: CRISPR Optimized MAGE Recombineering. *Sci. Rep.* 6, 19452.
- (117) 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.
- (118) 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.

- (119) Wingfield, P. T. (2015) Overview of the Purification of Recombinant Proteins. *Curr. Protoc. Protein Sci.* 80, 289–313.
- (120) Kumari, S., Beatty, C. M., Browning, D. F., Busby, S. J. W., Simel, E. J., and Hovel-Miner, G. (2000) Regulation of acetyl coenzyme A synthetase in escherichia coli. *J. Bacteriol.* 182, 4173–4179.
- (121) Raaijmakers, H. C. A., and Romão, M. J. (2006) Formate-reduced E. coli formate dehydrogenase H: The reinterpretation of the crystal structure suggests a new reaction mechanism. *J. Biol. Inorg. Chem.* 11, 849–854.
- (122) Waygood, E. B., Mort, J. S., and Sanwal, B. D. (1976) The Control of Pyruvate Kinase of Escherichia coli. Binding of Substrate and Allosteric Effectors to the Enzyme Activated by Fructose 1,6-Bisphosphate. *Biochemistry* 15, 277–282.
- (123) Baez, M., Cabrera, R., Pereira, H. M., Blanco, A., Villalobos, P., Ramírez-Sarmiento, C. A., Caniuguir, A., Guixé, V., Garratt, R. C., and Babul, J. (2013) A Ribokinase family conserved monovalent cation binding site enhances the MgATP-induced inhibition in E. coli phosphofructokinase-2. *Biophys. J.* 105, 185–193.
- (124) Shen, L. C., and Atkinson, D. E. (1970) Regulation of Pyruvate Dehydrogenase from Escherichia coli INTERACTIONS OF ADENYLATE ENERGY CHARGE AND OTHER REGULATORY PARAMETERS*. *Biol. Chem.* 246, 6974–6978.
- (125) Zhu, Y., Eiteman, M. A., Altman, R., and Altman, E. (2008) High glycolytic flux improves pyruvate production by a metabolically engineered Escherichia coli strain. *Appl. Environ. Microbiol.* 74, 6649–6655.
- (126) Aebersold, R., Burlingame, A. L., and Bradshaw, R. A. (2013, September) Western blots versus selected reaction monitoring assays: Time to turn the tables? *Mol. Cell. Proteomics*. Mol Cell Proteomics.
- (127) Jayasena, T., Poljak, A., Braid, N., Zhong, L., Rowlands, B., Muenchhoff, J., Grant, R., Smythe, G., Teo, C., Raftery, M., and Sachdev, P. (2016) Application of Targeted Mass

Spectrometry for the Quantification of Sirtuins in the Central Nervous System. *Sci. Rep.* 6.

(128) Kim, S. R., Nguyen, T. V., Seo, N. R., Jung, S., An, H. J., Mills, D. A., and Kim, J. H. (2015) Comparative proteomics: assessment of biological variability and dataset comparability. *BMC Bioinformatics* 16, 121.

(129) Kabir, M. M., Pei, Y. H., and Shimizu, K. (2005) Effect of IdhA 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.

(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) Weinstock, M. T., Heseck, E. D., Wilson, C. M., and Gibson, D. G. (2016) *Vibrio natriegens* as a fast-growing host for molecular biology 13.

(135) Hoffart, E., Grenz, S., Lange, J., Nitschel, R., Müller, F., Schwentner, A., Feith, A., Lenfers-lücker, M., and Takors, R. (2017) High Substrate Uptake Rates Empower *Vibrio natriegens* as Production Host for Industrial Biotechnology. *Appl. Environ. Microbiol.* 83, 1–10.

(136) Long, C. P., Gonzalez, J. E., Cipolla, R. M., and Antoniewicz, M. R. (2017) Metabolism of the fast-growing bacterium *Vibrio natriegens* elucidated by ¹³C metabolic flux analysis. *Metab. Eng.* 44, 191–197.

(137) 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.

(138) 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.

(139) Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J., and Tan, S. (2005) Comparison of affinity tags for protein purification. *Protein Expr. Purif.* 41, 98–105.

(140) 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.

(141) 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.

(142) 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.

(143) Clomburg, J. M., and Gonzalez, R. (2010) Biofuel production in *Escherichia coli*: the role of metabolic engineering and synthetic biology. *Appl. Microbiol. Biotechnol.* 2010 862 86, 419–434.

(144) Gustavsson, M., and Lee, S. Y. (2016) Prospects of microbial cell factories developed through systems metabolic engineering. *Microb. Biotechnol.* 9, 610–617.

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

(146) Hoekman, S. K., and Broch, A. (2018) Environmental implications of higher ethanol production and use in the U.S.: A literature review. Part II – Biodiversity, land use change, GHG emissions, and sustainability. *Renew. Sustain. Energy Rev.*

- (147) Garcia, D., Dinglasan, J. L., Shrestha, H., Abraham, P., Hettich, R., and Doktycz, M. (2020) A Lysate Proteome Engineering Strategy for Enhancing Cell-Free Metabolite Production. *bioRxiv* 2020.04.05.026393.
- (148) Chen, Y., Daviet, L., Schalk, M., Siewers, V., and Nielsen, J. (2013) Establishing a platform cell factory through engineering of yeast acetyl-CoA metabolism. *Metab. Eng.* 15.
- (149) Hollinshead, W., He, L., and Tang, Y. J. (2014) Biofuel production: An odyssey from metabolic engineering to fermentation scale-up. *Front. Microbiol.* 5, 344.
- (150) Seok, J., Ko, Y. J., Lee, M. E., Hyeon, J. E., and Han, S. O. (2019) Systems metabolic engineering of *Corynebacterium glutamicum* for the bioproduction of biliverdin via protoporphyrin independent pathway. *J. Biol. Eng.* 13, 1–13.
- (151) He, L., Xiao, Y., Gebreselassie, N., Zhang, F., Antoniewicz, M. R., Tang, Y. J., and Peng, L. (2014) Central Metabolic Responses to the Overproduction of Fatty Acids in *Escherichia coli* Based on ¹³C-Metabolic Flux Analysis. *Biotechnol. Bioeng.* 111, 575.
- (152) Alter, T. B., Blank, L. M., and Ebert, B. E. (2021) Proteome Regulation Patterns Determine *Escherichia coli* Wild-Type and Mutant Phenotypes. *mSystems* 6.
- (153) Hunt, J. P., Barnett, R. J., Robinson, H., Soltani, M., Nelson, J. A. D., and Bundy, B. C. (2021) Rapid sensing of clinically relevant glutamine concentrations in human serum with metabolically engineered *E. coli*-based cell-free protein synthesis. *J. Biotechnol.* 325, 389–394.
- (154) Dinglasan, J. L. N., Reeves, D. T., Hettich, R. L., and Doktycz, M. J. (2021) Liquid chromatography coupled to refractive index or mass spectrometric detection for metabolite profiling in lysate-based cell-free systems. *J. Vis. Exp.* 2021.
- (155) Kim, Y., Ingram, L. O., and Shanmugam, K. T. (2007) Construction of an *Escherichia coli* K-12 Mutant for Homoethanologenic Fermentation of Glucose or Xylose without Foreign Genes. *Appl. Environ. Microbiol.* 73, 1766.
- (156) Andorfer, M. C., Backman, L. R. F., Li, P. L., Ulrich, E. C., and Drennan, C. L. (2021) Rescuing activity of oxygen-damaged pyruvate formate-lyase by a spare part protein. *J. Biol.*

Chem. 297, 101423.

(157) Zhang, Y., Dai, Z., Krivoruchko, A., Chen, Y., Siewers, V., and Nielsen, J. (2015) Functional pyruvate formate lyase pathway expressed with two different electron donors in *Saccharomyces cerevisiae* at aerobic growth. *FEMS Yeast Res.* 15.

(158) Dopp, J. L., Jo, Y. R., and Reuel, N. F. (2019) Methods to reduce variability in E. Coli-based cell-free protein expression experiments. *Synth. Syst. Biotechnol.* 4, 204–211.

(159) Zhu, J., and Shimizu, K. (2005) Effect of a single-gene knockout on the metabolic regulation in *Escherichia coli* for D-lactate production under microaerobic condition. *Metab. Eng.* 7, 104–115.

(160) 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.

(161) Waygood, E. B., and Sanwal, B. D. (1974) The Control of Pyruvate Kinases of *Escherichia coli*. *J. Biol. Chem.* 249.

(162) Cassey, B., Guest, J. R., and Attwood, M. M. (1998) Environmental control of pyruvate dehydrogenase complex expression in *Escherichia coli*. *FEMS Microbiol. Lett.* 159.

(163) Kessler, D., Leibrecht, I., and Knappe, J. (1991) Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase activities of *Escherichia coli* reside on a polymeric protein particle encoded by *adhE*. *FEBS Lett.* 281, 59–63.

(164) Jewett, M. C., Calhoun, K. A., Voloshin, A., Wu, J. J., and Swartz, J. R. (2008) An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol. Syst. Biol.* 4, 220.

(165) Garenne, D., Haines, M. C., Romantseva, E. F., Freemont, P., Strychalski, E. A., and Noireaux, V. (2021) Cell-free gene expression. *Nat. Rev. Methods Prim.* 2021 11 1, 1–18.

(166) Förster, A. H., and Gescher, J. (2014) Metabolic Engineering of *Escherichia coli* for Production of Mixed-Acid Fermentation End Products. *Front. Bioeng. Biotechnol.* 2.

- (167) Clark, D. P. (1989) The fermentation pathways of *Escherichia coli*. *FEMS Microbiol. Lett.* 63.
- (168) Leonardo, M. R., Dailly, Y., and Clark, D. P. (1996) Role of NAD in regulating the *adhE* gene of *Escherichia coli*. *J. Bacteriol.* 178, 6013–6018.
- (169) Tamiev, B. D., Dopp, J. L., and Reuel, N. F. (2021) Anaerobic conditioning of *E. Coli* Cell Lysate for Enhanced In vitro Protein Synthesis. *ACS Synth. Biol.* 10, 716.
- (170) Calhoun, K. A., and Swartz, J. R. (2005) Energizing cell-free protein synthesis with glucose metabolism. *Biotechnol. Bioeng.* 90, 606–613.
- (171) Wilks, J. C., and Slonczewski, J. L. (2007) pH of the cytoplasm and periplasm of *Escherichia coli*: Rapid measurement by green fluorescent protein fluorimetry. *J. Bacteriol.* 189, 5601–5607.
- (172) Ugurbil, K., Rottenberg, H., Glynn, P., and Shulman, R. G. (1978) Biochemistry 1p nuclear magnetic resonance studies of bioenergetics and glycolysis in anaerobic *Escherichia coli* cells (ApH/adenosine triphosphatase/dicyclohexylcarbodiimide) 75, 2244–2248.
- (173) Lund, P., Tramonti, A., De Biase, D., Pasteur-Fondazione, I., and Bolognetti, C. (2014) Coping with low pH: molecular strategies in neutralophilic bacteria. *FEMS Microbiol. Rev.* 38, 1091–1125.
- (174) Carteni, F., Occhicone, A., Giannino, F., Vincenot, C. E., de Alteriis, E., Palomba, E., and Mazzoleni, S. (2020) A General Process-Based Model for Describing the Metabolic Shift in Microbial Cell Cultures. *Front. Microbiol.* 11, 2332.
- (175) Basan, M., Hui, S., and Williamson, J. R. (2017) ArcA overexpression induces fermentation and results in enhanced growth rates of *E. coli*. *Sci. Reports* 2017 71 7, 1–7.
- (176) Membrillo-Hernández, J., and Lin, E. C. C. (1999) Regulation of Expression of the *adhE* Gene, Encoding Ethanol Oxidoreductase in *Escherichia coli*: Transcription from a Downstream Promoter and Regulation by Fnr and RpoS. *J. Bacteriol.* 181, 7571.
- (177) Yasid, N. A., Rolfe, M. D., Green, J., and Williamson, M. P. (2016) Homeostasis of metabolites in *Escherichia coli* on transition from anaerobic to aerobic conditions and the transient

secretion of pyruvate. *R. Soc. Open Sci.* 3.

(178) Fujiwara, K., and Nomura, S.-I. M. (2013) Condensation of an Additive-Free Cell Extract to Mimic the Conditions of Live Cells.

(179) Takahashi, K., Sato, G., Doi, N., and Fujiwara, K. (2021) A Relationship between NTP and Cell Extract Concentration for Cell-Free Protein Expression. *Life* 11.

(180) Scopes, R. K., and Griffiths-Smith, K. (1986) FERMENTATION CAPABILITIES OF *Zymomonas mobilis* GLYCOLYTIC ENZYMES. *Biotechnol. Lett.* 8, 53–656.

(181) Kimple, M. E., Brill, A. L., and Pasker, R. L. (2013) Overview of Affinity Tags for Protein Purification. *Curr. Protoc. Protein Sci.* 73, Unit.

(182) Islam, M. R., Kwak, J. W., Lee, J. soo, Hong, S. W., Khan, M. R. I., Lee, Y., Lee, Y., Lee, S. W., and Hwang, I. (2019) Cost-effective production of tag-less recombinant protein in *Nicotiana benthamiana*. *Plant Biotechnol. J.* 17, 1094.

(183) Wang, D., and Hong, J. (2014) Purification of a recombinant protein with cellulose-binding module 3 as the affinity tag. *Methods Mol. Biol.* 1177, 35–45.

(184) Yin, G., Stephenson, H. T., Yang, J., Li, X., Armstrong, S. M., Heibeck, T. H., Tran, C., Masikat, M. R., Zhou, S., Stafford, R. L., Yam, A. Y., Lee, J., Steiner, A. R., Gill, A., Penta, K., Pollitt, S., Baliga, R., Murray, C. J., Thanos, C. D., Mcevoy, L. M., Sato, A. K., and Hallam, T. J. RF1 attenuation enables efficient non-natural amino acid incorporation for production of homogeneous antibody drug conjugates OPEN.

(185) Baneyx, F., and Georgiou, G. (1990) In vivo degradation of secreted fusion proteins by the *Escherichia coli* outer membrane protease OmpT. *J. Bacteriol.* 172, 491.

(186) Gottesman, S. (1996) PROTEASES AND THEIR TARGETS IN *ESCHERICHIA COLI* 1. *Annu. Rev. Genet* 30, 465–506.

(187) Martínez-Núñez, M. A., and López, V. E. L. y. (2016) Nonribosomal peptides synthetases and their applications in industry. *Sustain. Chem. Process.* 4, 13.

- (188) Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L., and Sivonen, K. (2014) Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9259–9264.
- (189) Felnagle, E. A., Jackson, E. E., Chan, Y. A., Podevels, A. M., Berti, A. D., McMahon, M. D., and Thomas, M. G. (2008) Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products. *Mol. Pharm.* **5**, 191.
- (190) Miller, B. R., and Gulick, A. M. (2016) Structural Biology of Nonribosomal Peptide Synthetases. *Methods Mol. Biol.* **1401**, 3–29.
- (191) Ishikawa, F., Nohara, M., Nakamura, S., Nakanishi, I., and Tanabe, G. (2020) Precise Probing of Residue Roles by NRPS Code Swapping: Mutation, Enzymatic Characterization, Modeling, and Substrate Promiscuity of Aryl Acid Adenylation Domains. *Biochemistry* **59**, 351–363.
- (192) Bian, X., Plaza, A., Yan, F., Zhang, Y., and Müller, R. (2015) Rational and efficient site-directed mutagenesis of adenylation domain alters relative yields of luminide derivatives in vivo. *Biotechnol. Bioeng.* **112**, 1343–1353.
- (193) Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505.
- (194) Calcott, M. J., Owen, J. G., and Ackerley, D. F. (2020) Efficient rational modification of non-ribosomal peptides by adenylation domain substitution. *Nat. Commun.* **2020 111 11**, 1–10.
- (195) Bozhüyük, K. A. J., Fleischhacker, F., Linck, A., Wesche, F., Tietze, A., Niesert, C. P., and Bode, H. B. (2017) De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **2017 103 10**, 275–281.
- (196) Kries, H., Niquille, D. L., and Hilvert, D. (2015) A subdomain swap strategy for reengineering nonribosomal peptides. *Chem. Biol.* **22**, 640–648.
- (197) Chiocchini, C., Linne, U., and Stachelhaus, T. (2006) In Vivo Biocombinatorial Synthesis of Lipopeptides by COM Domain-Mediated Reprogramming of the Surfactin Biosynthetic Complex.

Chem. Biol. **13**, 899–908.

(198) Beer, R., Herbst, K., Ignatiadis, N., Kats, I., Adlung, L., Meyer, H., Niopek, D., Christiansen, T., Georgi, F., Kurzawa, N., Meichsner, J., Rabe, S., Riedel, A., Sachs, J., Schessner, J., Schmidt, F., Walch, P., Niopek, K., Heinemann, T., Eils, R., and Di Ventura, B. (2014) Creating functional engineered variants of the single-module non-ribosomal peptide synthetase IndC by T domain exchange. *Mol. Biosyst.* **10**, 1709–1718.

(199) Kries, H., Wachtel, R., Pabst, A., Wanner, B., Niquille, D., and Hilvert, D. (2014) Reprogramming Nonribosomal Peptide Synthetases for “Clickable” Amino Acids. *Angew. Chemie Int. Ed.* **53**, 10105–10108.

(200) Bozhüyük, K. A. J., Linck, A., Tietze, A., Kranz, J., Wesche, F., Nowak, S., Fleischhacker, F., Shi, Y. N., Grün, P., and Bode, H. B. (2019) Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **2019 117 11**, 653–661.

(201) Thirlway, J., Lewis, R., Nunns, L., Al Nakeeb, M., Styles, M., Struck, A. W., Smith, C. P., and Micklefield, J. (2012) Introduction of a Non-Natural Amino Acid into a Nonribosomal Peptide Antibiotic by Modification of Adenylation Domain Specificity. *Angew. Chemie Int. Ed.* **51**, 7181–7184.

(202) Kudo, F., Miyanaga, A., and Eguchi, T. (2019) Structural basis of the nonribosomal codes for nonproteinogenic amino acid selective adenylation enzymes in the biosynthesis of natural products. *J. Ind. Microbiol. Biotechnol.* **46**, 515–536.

(203) Villiers, B., and Hollfelder, F. (2011) Directed evolution of a gatekeeper domain in nonribosomal peptide synthesis. *Chem. Biol.* **18**, 1290–1299.

(204) Challis, G. L., and Naismith, J. H. (2004) Structural aspects of non-ribosomal peptide biosynthesis. *Curr. Opin. Struct. Biol.* **14**, 748–756.

(205) Li, J., Zhang, L., and Liu, W. (2018) Cell-free synthetic biology for in vitro biosynthesis of pharmaceutical natural products. *Synth. Syst. Biotechnol.* **3**, 83.

- (206) Li, J., Zhang, L., and Liu, W. (2018, June 1) Cell-free synthetic biology for in vitro biosynthesis of pharmaceutical natural products. *Synth. Syst. Biotechnol.* KeAi Communications Co.
- (207) Siebels, I., Nowak, S., Heil, C. S., Tufar, P., Cortina, N. S., Bode, H. B., and Grininger, M. (2020) Cell-Free Synthesis of Natural Compounds from Genomic DNA of Biosynthetic Gene Clusters. *ACS Synth. Biol.* 9, 2418–2426.
- (208) Goering, A. W., Li, J., McClure, R. A., Thomson, R. J., Jewett, M. C., and Kelleher, N. L. (2017) In vitro reconstruction of nonribosomal peptide biosynthesis directly from DNA using cell-free protein synthesis. *ACS Synth. Biol.* 6, 39–44.
- (209) Jewett, M. C., and Swartz, J. R. (2004) Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol. Bioeng.* 86, 19–26.
- (210) Vilkhovoy, M., Horvath, N., Shih, C. H., Wayman, J. A., Calhoun, K., Swartz, J., and Varner, J. D. (2018) Sequence Specific Modeling of E. Coli Cell-Free Protein Synthesis. *ACS Synth. Biol.* 7, 1844–1857.
- (211) Goering, A. W., Li, J., McClure, R. A., Thomson, R. J., Jewett, M. C., and Kelleher, N. L. (2017) In vitro reconstruction of nonribosomal peptide biosynthesis directly from DNA using cell-free protein synthesis. *ACS Synth. Biol.* 6, 39–44.
- (212) 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.* 89, 11443–11451.
- (213) Wick, S., Walsh, D. I., Bobrow, J., Hamad-Schifferli, K., Kong, D. S., Thorsen, T., Mrozczyk, K., and Carr, P. A. (2019) PERSIA for Direct Fluorescence Measurements of Transcription, Translation, and Enzyme Activity in Cell-Free Systems. *ACS Synth. Biol.* 8, 1010–1025.

- (214) Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., and Khosla, C. (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* (80-.). 291, 1790–1792.
- (215) Li, J., Wang, H., Kwon, Y.-C., and Jewett, M. C. (2017) Establishing a high yielding *streptomyces* -based cell-free protein synthesis system. *Biotechnol. Bioeng.* 114, 1343–1353.
- (216) Li, J., Wang, H., and Jewett, M. C. (2018) Expanding the palette of *Streptomyces*-based cell-free protein synthesis systems with enhanced yields. *Biochem. Eng. J.* 130, 29–33.
- (217) Moore, S. J., Lai, H. E., Needham, H., Polizzi, K. M., and Freemont, P. S. (2017) *Streptomyces venezuelae* TX-TL - a next generation cell-free synthetic biology tool. *Biotechnol. J.* 12.
- (218) Moore, S. J., Lai, H. E., Chee, S. M., Toh, M., Coode, S., Chengan, K., Capel, P., Corre, C., De Los Santos, E. L. C., and Freemont, P. S. (2021) A *Streptomyces venezuelae* Cell-Free Toolkit for Synthetic Biology. *ACS Synth. Biol.* 10, 402–411.
- (219) Kelwick, R., Webb, A. J., MacDonald, J. T., and Freemont, P. S. (2016) Development of a *Bacillus subtilis* cell-free transcription-translation system for prototyping regulatory elements. *Metab. Eng.* 38, 370–381.
- (220) Pang, B., Pang, B., Pang, B., Chen, Y., Chen, Y., Gan, F., Gan, F., Gan, F., Yan, C., Jin, L., Gin, J. W., Gin, J. W., Petzold, C. J., Petzold, C. J., Keasling, J. D., Keasling, J. D., Keasling, J. D., Keasling, J. D., Keasling, J. D., and Keasling, J. D. (2020) Investigation of Indigoidine Synthetase Reveals a Conserved Active-Site Base Residue of Nonribosomal Peptide Synthetase Oxidases. *J. Am. Chem. Soc.* 142, 10931–10935.
- (221) Owen, J. G., Copp, J. N., and Ackerley, D. F. (2011) Rapid and flexible biochemical assays for evaluating 4'- phosphopantetheinyl transferase activity. *Biochem. J.* 436, 709–717.
- (222) Owen, J. G., Calcott, M. J., Robins, K. J., and Ackerley, D. F. (2016) Generating Functional Recombinant NRPS Enzymes in the Laboratory Setting via Peptidyl Carrier Protein Engineering. *Cell Chem. Biol.* 23, 1395–1406.

- (223) Hoffmann, C., Gaietta, G., Zürn, A., Adams, S. R., Terrillon, S., Ellisman, M. H., Tsien, R. Y., and Lohse, M. J. (2010) Fluorescent labeling of tetracysteine-tagged proteins in intact cells. *Nat. Protoc.* 2010 510 5, 1666–1677.
- (224) Stevens, D. C., Conway, K. R., Pearce, N., Villegas-Peñaranda, L. R., Garza, A. G., and Boddy, C. N. (2013) Alternative Sigma Factor Over-Expression Enables Heterologous Expression of a Type II Polyketide Biosynthetic Pathway in Escherichia coli. *PLoS One* 8.
- (225) Hillson, N. J., Rosengarten, R. D., and Keasling, J. D. (2012) J5 DNA assembly design automation software. *ACS Synth. Biol.* 1, 14–21.
- (226) Wehrs, M., Gladden, J. M., Liu, Y., Platz, L., Prah, J. P., Moon, J., Papa, G., Sundstrom, E., Geiselman, G. M., Tanjore, D., Keasling, J. D., Pray, T. R., Simmons, B. A., and Mukhopadhyay, A. (2019) Sustainable bioproduction of the blue pigment indigoidine: Expanding the range of heterologous products in: R. toruloides to include non-ribosomal peptides. *Green Chem.* 21, 3394–3406.
- (227) Douglas S. Auld, P. D., Peter A. Coassin, B. S., Nathan P. Coussens, P. D., Hensley, P., Klumpp-Thomas, C., Michael, S., G. Sitta Sittampalam, P. D., O. Joseph Trask, B. S., Bridget K. Wagner, P. D., Jeffrey R. Weidner, P. D., Mary Jo Wildey, P. D., and Jayme L. Dahlin, M. D. . P. D. (2020) Microplate Selection and Recommended Practices in High-throughput Screening and Quantitative Biology. *Assay Guid. Man.*
- (228) Liang, Y., Woodle, S. A., Shibeko, A. M., Lee, T. K., and Ovanesov, M. V. (2013) Correction of microplate location effects improves performance of the thrombin generation test. *Thromb. J.* 11.
- (229) Kim, D. M., and Swartz, J. R. (1999) Prolonging cell-free protein synthesis with a novel ATP regeneration system. *Biotechnol. Bioeng.* 66, 180–188.
- (230) Yang, C. ;, Yang, M. ;, Zhao, W. ;, Ding, Y. ;, Wang, Y. ;, Li, J., Yang, C., Yang, M., Zhao, W., Ding, Y., Wang, Y., and Li, J. (2022) Establishing a Klebsiella pneumoniae-Based Cell-Free Protein Synthesis System. *Mol.* 2022, Vol. 27, Page 4684 27, 4684.

- (231) Des Soye, B. J., Davidson, S. R., Weinstock, M. T., Gibson, D. G., and Jewett, M. C. (2018) Establishing a High-Yielding Cell-Free Protein Synthesis Platform Derived from *Vibrio natriegens*. *ACS Synth. Biol.* 7, 2245–2255.
- (232) Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* 16, 4174–4183.
- (233) Guo, X., Zhu, Y., Bai, L., and Yang, D. (2020) The Protection Role of Magnesium Ions on Coupled Transcription and Translation in Lyophilized Cell-Free System. *ACS Synth. Biol.* 9, 856–863.
- (234) Failmezger, J., Scholz, S., Blombach, B., and Siemann-Herzberg, M. (2018) Cell-Free Protein Synthesis From Fast-Growing *Vibrio natriegens*. *Front. Microbiol.* 9, 1146.
- (235) Kim, J., Copeland, C. E., Seki, K., Vögeli, B., and Kwon, Y. C. (2020) Tuning the Cell-Free Protein Synthesis System for Biomanufacturing of Monomeric Human Filaggrin. *Front. Bioeng. Biotechnol.* 8.
- (236) Kim, D.-M., and Swartz, J. R. (2000) Prolonging Cell-Free Protein Synthesis by Selective Reagent Additions.
- (237) Whittaker, J. W. (2013) Cell-free protein synthesis: the state of the art. *Biotechnol. Lett.* 35, 143.
- (238) Voloshin, A. M., and Swartz, J. R. (2008) Large-Scale Batch Reactions for Cell-Free Protein Synthesis, in *Cell-Free Protein Synthesis: Methods and Protocols* (Voloshin, A. M., and Swartz, J. R., Eds.) 1st ed., pp 207–235. Wiley-VCH Verlag GmbH & Co.KGaA, Weinheim, Germany.
- (239) Kim, D.-M., and Swartz, J. R. (2001) Regeneration of Adenosine Triphosphate from Glycolytic Intermediates for Cell-Free Protein Synthesis.
- (240) Takahashi, H., Kumagai, T., Kitani, K., Mori, M., Matoba, Y., and Sugiyama, M. (2007) Cloning and characterization of a *Streptomyces* single module type non-ribosomal peptide synthetase catalyzing a blue pigment synthesis. *J. Biol. Chem.* 282, 9073–9081.

- (241) Calhoun, K. A., and Swartz, J. R. (2006) Total amino acid stabilization during cell-free protein synthesis reactions. *J. Biotechnol.* 123, 193–203.
- (242) Hahn, M., and Stachelhaus, T. (2004) Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15585.
- (243) Klapper, M., Braga, D., Lackner, G., Herbst, R., and Stallforth, P. (2018) Bacterial Alkaloid Biosynthesis: Structural Diversity via a Minimalistic Nonribosomal Peptide Synthetase. *Cell Chem. Biol.* 25, 659-665.e9.
- (244) Gulick, A. M., and Aldrich, C. C. Trapping Interactions Between Catalytic Domains and Carrier Proteins of Modular Biosynthetic Enzymes with Chemical Probes.
- (245) Bozhüyük, K. A., Micklefield, J., and Wilkinson, B. (2019) Engineering enzymatic assembly lines to produce new antibiotics. *Curr. Opin. Microbiol.* 51, 88–96.
- (246) Gruenewald, S., Mootz, H. D., Stehmeier, P., and Stachelhaus, T. (2004) In Vivo Production of Artificial Nonribosomal Peptide Products in the Heterologous Host *Escherichia coli*. *Appl. Environ. Microbiol.* 70, 3282.
- (247) Wannier, T. M., Nyerges, A., Kuchwara, H. M., Czikkely, M., Balogh, D., Filsinger, G. T., Borders, N. C., Gregg, C. J., Lajoie, M. J., Rios, X., Pál, C., and Church, G. M. (2020) Improved bacterial recombineering by parallelized protein discovery. *Proc. Natl. Acad. Sci. U. S. A.* 117, 13689–13698.
- (248) Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J. L., Makarova, K. S., Koonin, E. V., and Zhang, F. (2019) RNA-guided DNA insertion with CRISPR-associated transposases. *Science* (80-.). 364, 48–53.
- (249) Rolf, J., Rosenthal, K., and Lütz, S. (2019) Application of Cell-Free Protein Synthesis for Faster Biocatalyst Development. *Catal.* 2019, Vol. 9, Page 190 9, 190.
- (250) Yuan, Q., Wu, M., Liao, Y., Liang, S., Lu, Y., and Lin, Y. (2023) Rapid prototyping enzyme homologs to improve titer of nicotinamide mononucleotide using a strategy combining cell-free

protein synthesis with split GFP. *Biotechnol. Bioeng.*

VITA

Jaime Lorenzo Naval Dinglasan was born in Manila, Philippines to Dr. Jaime Lopez Dinglasan and Dr. Carmen Isabel Naval Dinglasan. He was raised in Mandaluyong City, where he attended La Salle Green Hills High School. He received his Bachelor of Science in Biology from the University of the Philippines Diliman (UPD), Quezon City in 2017. As an undergraduate, he conducted his senior thesis with a focus on Bioinformatics and Molecular Biology under the supervision of Dr. Michael C. Velarde. He started his scientific career as a Research Assistant at UPD's Institute of Biology immediately after graduating. A year later, he pursued his doctorate in Genome Science and Technology at the University of Tennessee Knoxville (UTK) and Oak Ridge National Laboratory (ORNL). His doctoral research on Cell-free Synthetic Biology was performed with Dr. Mitchel J. Doktycz in the Molecular and Cellular Imaging Group at ORNL.