# INTERFACING SYSTEMS AND SYNTHETIC BIOLOGY FOR ADVANCEMENTS IN BACTERIAL BIOSENSOR ENGINEERING

A Dissertation
Presented to
The Academic Faculty

By

April M. Miguez

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in
Bioengineering

Georgia Institute of Technology

May 2021

# INTERFACING SYSTEMS AND SYNTHETIC BIOLOGY FOR ADVANCEMENTS IN BACTERIAL BIOSENSOR ENGINEERING

Approved by:

Dr. Mark P. Styczynski, Advisor
School of Chemical & Biomolecular Engineering
*Georgia Institute of Technology*

Dr. Hang Lu
School of Chemical & Biomolecular Engineering
*Georgia Institute of Technology*

Dr. Lily S. Cheung
School of Chemical & Biomolecular Engineering
*Georgia Institute of Technology*

Dr. Ronghu Wu
School of Chemistry & Biochemistry
*Georgia Institute of Technology*

Dr. Meleah A. Hickman
Department of Biology
*Emory University*

Date Approved: April 21, 2021

# ACKNOWLEDGEMENTS

This thesis would not have been possible without the many people who helped me along the way.

I would like to express my deepest appreciation to my advisor, Mark Styczynski, who has been instrumental in guiding me throughout my scientific endeavors during my PhD studies. Without his continuous support and guidance, this thesis would not have been possible. By challenging me to think critically and creatively in every step of the way, he has helped me become a better scientist and researcher.

I am also extremely grateful to my thesis committee members – Dr. Cheung, Dr. Hickman, Dr. Lu, and Dr. Wu – who have given me excellent advice and insights that have shaped this thesis and helped me to think more critically about my research.

I would also like to thank the entire Styczynski group for their stimulating discussions, constant motivation, and fruitful collaborations. I would especially like to thank Monica McNerney, Yan Zhang, and Fernanda Piorino without whom this work would have not been accomplished.

Many thanks to my colleagues at Georgia Tech for their endless support throughout my time in graduate school.

Finally, my success would not have been possible without my parents Bryan and June and my siblings for their continuous love and encouragement throughout my entire academic journey, and my fiancé Enio and his family for being my home away from home.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF SYMBOLS AND ABBREVIATIONS

| | |
|---|---|
| ADP | Adenosine diphosphate |
| ANOVA | Analysis of variance |
| ANOVA2 | Two-way analysis of variance |
| ATP | Adenosine triphosphate |
| Cas | CRISPR-associated protein |
| CFE | Cell-free expression |
| CoA | Coenzyme A |
| CRISPR | Clustered regularly interspersed short palindromic repeats |
| CRISPRi | CRISPR interference |
| CTP | Cytidine triphosphate |
| dCas | Catalytically dead Cas |
| DHAP | Dihydroxyacetone phosphate |
| DMSO | Dimethyl sulfoxide |
| DNA | Deoxyribonucleic acid |
| DTT | Dithiothreitol |
| EDTA | Ethylenediaminetetraacetic acid |
| FAMES | Fatty acid methyl esters |
| FDR | False discovery rate |
| FPP | Farnesyl pyrophosphate |
| G6P | Glucose-6-phosphate |
| GABA | Gamma-aminobutyric acid |
| GC | Gas chromatography |

| | |
|---|---|
| GCxGC | 2-Dimensional gas chromatography |
| GD | Grown with glucose and dialyzed |
| GFP | Green fluorescent protein |
| GN | Grown with glucose and not dialyzed |
| gRNA | Guide RNA |
| GTP | Guanosine triphosphate |
| HMG-CoA | 3-Hydroxy-3-methylglutaryl coenzyme A |
| HPLC | High performance liquid chromatography |
| IPP | Isopentenyl pyrophosphate |
| KNN | K-nearest-neighbors |
| LB | Luria broth |
| LC | Liquid chromatography |
| m/z | Mass to charge ratio |
| MS | Mass spectrometry |
| MSTFA | N-methyl-N-(trimethylsilyl) trifluoroacetamide |
| NAD | Nicotinamide adenine dinucleotide |
| NADH | Nicotinamide adenine dinucleotide hydride |
| ND | Grown without glucose and dialyzed |
| NMR | Nuclear magnetic resonance |
| NN | Grown without glucose and not dialyzed |
| OAA | Oxaloacetate |
| OD | Optical density |
| PC | Principal component |
| PCA | Principal component analysis |
| PEP | Phosphoenolpyruvate |

| | |
|---|---|
| PLS-DA | Partial least-squares discriminant analysis |
| PURE | Purified recombinant elements |
| QC | Quality control |
| RBS | Ribosomal binding site |
| RNA | Ribonucleic acid |
| TCA | Tricarboxylic acid |
| TMCS | Trimethylchlorosilane |
| TOF | Time-of-flight |
| tRNA | Transfer RNA |
| UTP | Uridine triphosphate |

# SUMMARY

Current detection platforms ranging from clinical diagnostics to environmental pollutant monitoring often require a time-intensive sample analysis process involving expensive equipment and highly-trained staff. This has led to growing demands for faster, less expensive, more user-friendly platforms. Bacteria have the potential to meet these needs, as they can serve as inexpensive, robust biosensors that can be engineered to detect target molecules while providing fast, easily measurable readouts; however, genetic engineering efforts can often incite metabolic changes that limit biosensing performance. Cell-free bacteria-based biosensors, which use a bacterial protein lysate to perform transcription and translation, can avoid many of the challenges of whole-cell biosensor development, but the uncharacterized metabolic activity in cell-free systems creates a new set of obstacles that must be addressed for effective biosensor design. In this work, I use metabolomics (the systems-scale study of small molecule intermediates involved in the chemical reactions within biological systems) to address these key challenges in whole-cell and cell-free systems to improve their development for biosensing applications. For whole-cell systems, I explore the metabolic effects linked to expression and optimization of a well-characterized biosensor reporter system. For cell-free systems, I characterize their endogenous, dynamic metabolic activity and explore the metabolic impacts of various system perturbations. For both platforms, I identify key metabolites that limit the utility of both whole-cell and cell-free systems and present strategies to address some of the limitations in each platform to facilitate improved biosensor engineering and ultimately broaden the reach of whole-cell and cell-free bacteria-based biosensors.

# CHAPTER 1. INTRODUCTION

Bacteria-based biosensors have significant potential to address the growing need for improved detection platforms. Although researchers have learned to harness and manipulate the natural sense-and-respond genetic machinery from bacteria, genetic manipulation of these organisms often results in downstream metabolic changes that can limit the utility and applicability of bacteria for real-world sensing applications. Cell-free biosensors can avoid many of the challenges faced in whole-cell biosensor design; however, the fact that we have no idea what kind of metabolic activity even exists in these systems creates an entirely new set of challenges that must be addressed to increase the generalizability of a cell-free biosensing platform. For both whole-cell and cell-free biosensors, understanding the metabolic response to and implications of engineering efforts will substantially improve biosensor development.

## 1.1 Whole-cell bacteria for biosensor development

Biosensors are biological tools that are used to detect the presence of molecules of interest and have the potential to expand and improve current diagnostics, drug discovery, food monitoring, and environmental control efforts[1]. Specifically, whole-cell bacterial biosensors, which consist of engineered bacteria for detection of target analytes, are extremely promising platforms due to their low cost, robust nature, scalability, and genetic tractability[2]. They also have the possibility to transform the field of rapid point-of-care diagnostic testing, which provides users with real-time results in an easy to operate, low-cost platform, as they can reduce the need for expensive equipment and highly-trained staff[3].

In general, bacterial biosensors consist of two main components: (1) machinery to sense and respond to input signals and (2) a system to convert the biological response to measurable signal outputs[4] (Figure 1). For a biosensor to sense and respond effectively to target molecules, the genetic sensing machinery used must be selective and sensitive to the target input, and the resulting signal output should be easily measurable over any background noise even at low signal levels[5]. Although natural and synthetic sensing systems exist that can be used for the development of biosensors, these systems typically require extensive tuning to create a high performing biosensor that can detect the input molecule over the desired concentration range[6].

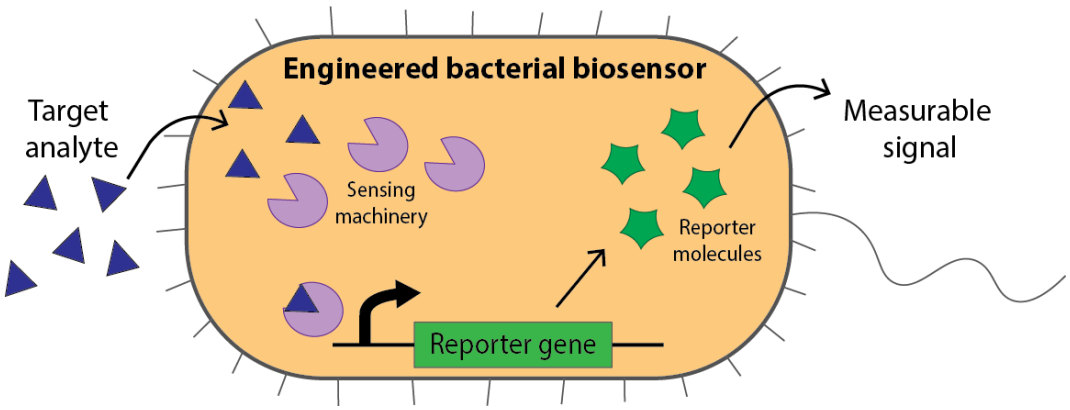


**Figure 1: Bacterial biosensor overview.**

Although advancements in synthetic biology (a field of research that aims to design and construct biological components to facilitate molecular engineering of cells) have led to the development of various tools that can be used to tune the dynamic range of biosensors, these engineering efforts can have undesirable effects on the metabolic state of the cell. Metabolite responsive transcription factors, two component systems, and regulatory RNAs are examples of effective synthetic biology tools that can be used to control the output ON- and OFF-states with minimal background noise[7]. However, these

approaches can adversely affect cellular metabolism and introduce cellular stress, which could negatively impact biosensor response, making tight control of metabolic state necessary to effectively control signal transduction. To regulate metabolic state, byproduct formation, environmental stress, and additional energetic burdens on the cell are minimized while the rate of output signal production in response to the presence of input is maximized[8]. Efforts to adjust transcription and translation rates, engineer enzymes, introduce precursor pathways, and express heterologous genes are examples of techniques that can be used to rebalance metabolism for improved biosensor response[9, 10].

The selection of the signal output known as the reporter is highly dependent on the specific needs of the biosensor application due to the unique advantages and limitations of each reporter system[11]. Fluorescent, bioluminescent, and colorimetric reporters are commonly used biosensor outputs[12, 13]. Fluorescent reporters such as green fluorescent protein (GFP) are stable, fast maturing outputs, but they require the use of specialized equipment for quantification under a specific light excitation. Bioluminescent reporters such as luciferase do not require an excitation light source but also requires specialized tools to measure the output. In contrast, colorimetric reporters such as β-galactosidase and crtA can often be qualitatively visualized by the naked eye. Although quantification of colorimetric output can be challenging without equipment[5], there exist many cases where qualitative results are sufficient, especially for point-of-care diagnostics.

Numerous bacteria-based biosensors have been developed using different reporter systems that sense and respond to various targets ranging from pathogens[14] and clinically-relevant biomarkers[15] to common pollutants[16, 17]. Yet major challenges remain that can limit the utility of whole-cell bacterial biosensors. Because whole-cell biosensors are

genetically engineered, they can experience severe stress due to plasmid maintenance, intra- and extracellular product toxicity, protein misfolding, and high gene expression that have downstream effects on metabolism. This cellular stress can redirect metabolic flux, creating a metabolic burden that can lead to decreased cell growth rate or terminal cell density, altered membrane composition, plasmid instability, and potentially even cell death[18]. Additionally, membrane permeability can also severely limit the detection range of these sensors[19].

## 1.2 Bacteria-based cell-free expression systems for biosensor development

Cell-free expression (CFE) systems have been shown to be a promising alternative to overcome these limitations in whole-cell systems[20-23]. CFE systems provide in vitro protein expression using extracted molecular machinery from cells, which can be collected from various types of organisms, including microbes, wheat germ, insect cells, and rabbit reticulocytes[24]. However, bacteria-based CFE systems are often used because they are the fastest to prepare, produce the highest titers of protein, and are the cheapest to manufacture[25, 26], with *Escherichia coli*-based CFE systems being the most popular. Bacterial CFE systems can be prepared as purified recombinant elements (PURE)[27-29] or crude protein extracts (lysates)[30, 31]. The PURE system uses purified, recombinantly-produced protein components involved in transcription and translation[32], whereas crude lysate systems use the cytosolic cellular material from bacteria that is separated from the membrane and genomic DNA.

Overall, both systems contain the necessary material to produce proteins, while lacking the typical whole-cell burdens such as cell division and membrane maintenance.

Their open-membrane platform (1) allows for the detection or production of molecules that cause cellular toxicity, (2) is easily monitored and manipulated, and (3) enables rapid prototyping[33]. They are even stable at room temperature after lyophilization and can be easily reactivated upon rehydration, expanding their applicability for field-friendly point-of-care diagnostics[22]. Although both the PURE and crude lysate systems can be used for biosensing applications, crude lysate systems can be favorable due to their low cost per reaction and the extensive efforts that have already been done to optimize lysate preparation methods and reaction conditions[34-37].

Bacterial crude lysate preparations typically follow the same overall steps: culturing cells, lysis, and lysate post-processing (Figure 2). Bacterial cells are typically cultured in rich medium such as LB or 2xYT and grown to log phase to optimize protein expression while minimizing the production of fermentation products[38]. At this stage, cells can be harvested, and their membranes can be disrupted without damaging intracellular proteins. Various approaches can be taken to lyse cells including homogenization[39], bead-beating[40], sonication[41], and lysozyme treatment[42]. After lysis, unwanted cellular debris and genomic DNA are separated from the crude lysate, which can then undergo further processing if desired, including incubation, dialysis, and centrifugation to further improve lysate productivity[39].

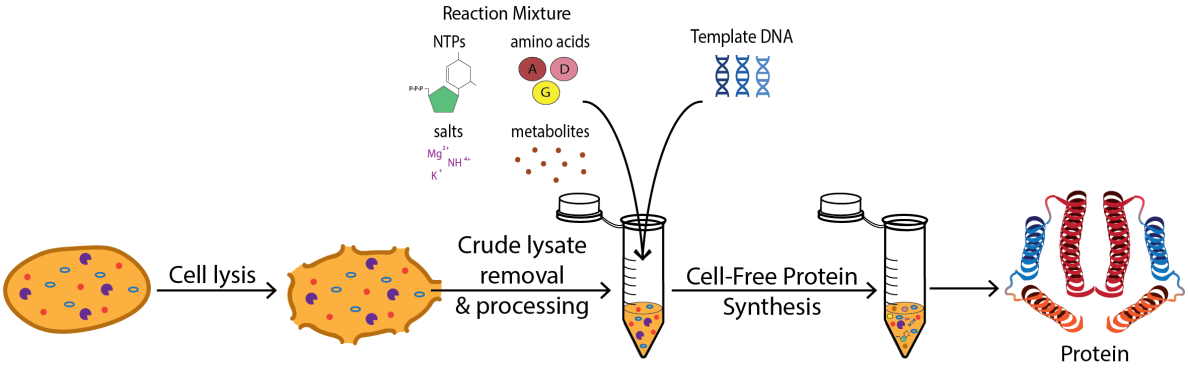


**Figure 2: Cell-free expression (CFE) systems overview.**
**CFE systems are prepared from bacterial cells. Cells are lysed, and the crude extract is extracted and processed. This bacterial lysate is combined with a reaction mixture and a DNA template to initiate transcription and translation of a target protein, which may itself be a product or may catalyze synthesis of a small molecule or another product. Figure from Miguez *et al*, *Industrial & Engineering Chemistry Research* 2019[43].**

Once the final lysate is collected, CFE reactions can be assembled. Complete CFE reactions consist of the protein lysate, a DNA template encoding a target protein or genetic circuit, and "reaction mixture." The reaction mixture contains energy sources, amino acids, cofactors, nucleoside triphosphates, tRNAs, substrates and salts. ATP is the primary energy source for the reaction, but molecules such as phosphoenolpyruvate, glucose-6-phosphate, or 3-phosphoglycerate are added to help regenerate ATP via substrate phosphorylation during the reaction[44].

Many of the recognition elements that have been applied to whole-cell biosensors such as toehold switches, aptamers, transcriptional regulators, and antibodies have also been successfully implemented in CFE sensors[19]. This has led to the development of CFE biosensors for a broad range of analytes ranging from nucleic acids to detect Ebola[45] and Zika viruses[46], small molecules such as quorum sensing molecules[11] and antibiotics[20], proteins such as alpha-fetoprotein[47], and metal ions such as zinc[23] and mercury[11].

Despite this progress, there are still hurdles in CFE systems that can preclude their use for biosensing purposes. Crude lysates are complex, undefined mixtures that are highly sensitive to small changes in lysate preparation method, which ultimately affect its activity in a CFE reaction. Batch-to-batch variability is a common issue resulting from this sensitivity to change that impacts the reproducibility[48], scalability[26], and standardization[49] of CFE systems. Limited reaction life is another prominent issue, which is thought to be a result of (1) depletion or degradation of essential metabolites for protein synthesis, (2) accumulation of small molecule transcriptional inhibitors and waste products, and/or (3) alteration of lysate pH[39, 50]. The buildup of unwanted molecules is thought to be in part due to the utilization of the substrates for ATP regeneration in the CFE reactions. Although these substrates have been shown to greatly improve the overall CFE reaction activity, they are thought to lead to an accumulation of inorganic phosphates, which could inhibit protein synthesis and reduce the translational life of the reaction[51]. Although efforts to mitigate these issues by exploring different energy sources has been successful at improving reaction yield[24], they have not completely resolved the issue of reaction longevity. Inactivation of expression machinery is not likely to be the cause because continuous CFE systems (which constantly replenish the reaction with fresh small-molecule reagents while removing reacted small-molecule products) have been able to extend reaction life; however, the exact small molecules that lead to this improvement are unknown, and batch (not continuous) CFE systems are preferable platforms for biosensing applications. Surprisingly, despite the many hypotheses suggesting metabolic-related causes for many of the issues in CFE systems, the metabolic profiles of these systems have yet to be characterized.

## 1.3 Bacterial biosensor development in our lab

Recently, our group has used both whole-cell and cell-free bacterial systems for the development of point-of-care diagnostic tools, facing the unique challenges of each system firsthand during the development of vitamin $B_{12}$ and zinc whole-cell and CFE biosensors. Despite the many advantages of CFE systems for biosensing design, we found for the development of a $B_{12}$ biosensor that whole-cell systems were preferable because they had improved sensitivity to $B_{12}$ due to their natural $B_{12}$ active membrane importers that concentrated the metabolite in the cell[52]. In contrast, for our zinc biosensor, CFE systems were preferable due to their ability to sensitively detect clinically relevant levels of zinc in a field- and user-friendly platform. However, the development for both types of sensor was limited by the need for extensive engineering efforts to tune their sensitivities and, specifically for the whole-cell sensors, to address toxicity related to the production of the metabolite pigment reporters[53]. Although our previous work highlights the utility of each platform for the detection of different molecules, we could more easily address issues and speed up our progress to create novel biosensors by having a deeper understanding of the exact metabolic impacts that our engineering efforts have on our biosensors.

## 1.4 Systems biology to address challenges in biosensor design

Systems biology, which aims to understand the complexity of biological systems by focusing on them as a whole rather than on individual components, is a promising approach to study the metabolic effects of genetic engineering on biosensors and has been proven to be extremely useful to advance microbial engineering[54]. Systems-level analyses typically use omics techniques that generate large data sets to provide a holistic

understanding of the different levels of regulation within a cell and can complement synthetic biology approaches for improved biosensor development[55, 56] (Figure 3). Transcriptomics, proteomics, and metabolomics are popular omics technologies that have been used to evaluate cellular physiology in response to microbial engineering. Transcriptomics and proteomics focus on studying the RNA transcripts and protein content with a cell, respectively, and can provide information on gene expression, regulation, and protein interactions that may be impacted by introducing recombinant genetic material[57]. Although these omics techniques provide in-direct measurements of metabolism, the changes in transcript- and protein-levels do not directly correspond to changes in metabolism[58].

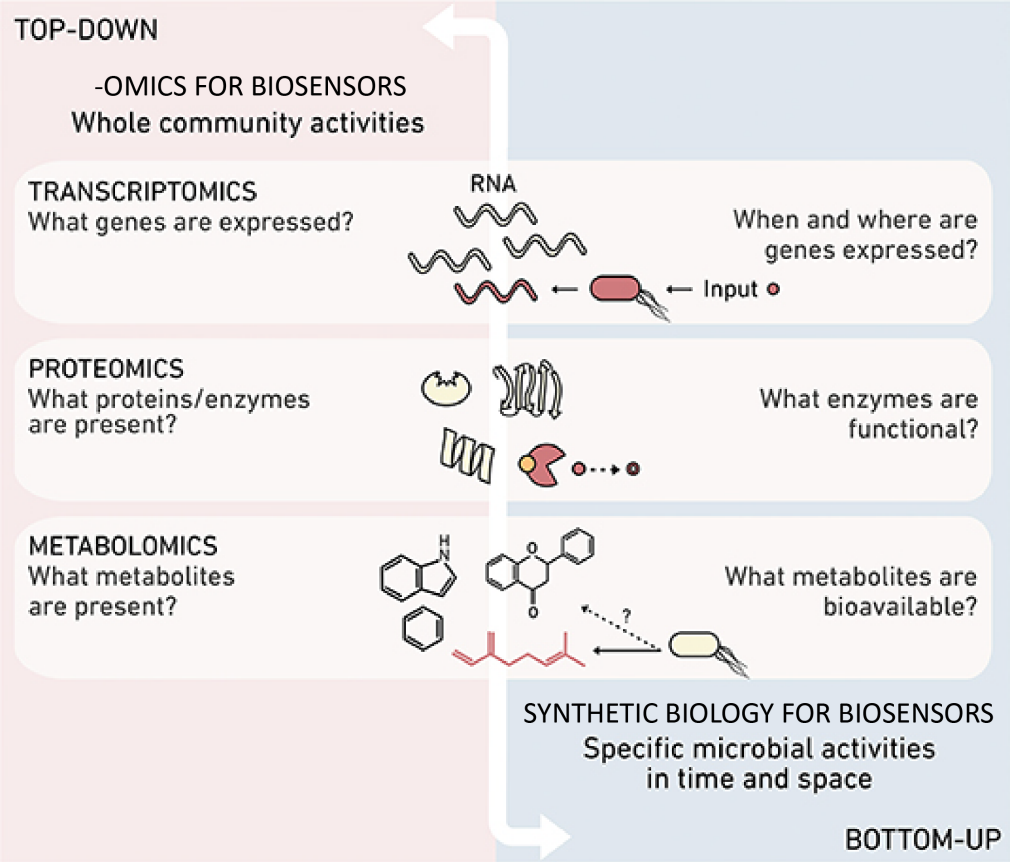


**Figure 3: Systems biology techniques for improved biosensor development. Omics techniques and synthetic biology take top-down and bottom-up approaches, respectively. Each provide unique and complementary perspectives on biosensing systems that can be used to inform optimization efforts. Figure modified from Del Valle *et al, Frontiers in Microbiology* 2021[56].**

To directly characterize the metabolite-level changes in biosensors, metabolomics can be used. Metabolomics is the study of the small-molecule intermediates of metabolism, which is defined as the set of chemical reactions in an organism required to provide energy and build cellular material[58]. Because metabolomics synthesizes the effects of genetic, transcriptional, and protein levels of regulation, it provides a downstream, endpoint readout of cellular metabolic activity. Although metabolomics has not been widely applied to biosensor engineering, it has become a valuable resource to inform genetic engineering of bacteria for improved production of valuable chemicals, such as fatty acids[59], butanol[60, 61],

and terpenoids[62, 63], highlighting the potential of metabolomics to inform engineering efforts for improved bacterial biosensor design. There has also been a recent growth of systems-level approaches to better understand the complex, undefined crude lysate used in CFE systems. Proteomics approaches in particular have been used by researchers to uncover the types of proteins present in the undefined, crude protein lysate[64-66], but there have been no metabolomics studies of these systems before my work and only a handful now[67-69]. With a better understanding of the endogenous metabolism within CFE systems, we would be able to better engineer biosensors.

## 1.5 Metabolomics analytical techniques

Efforts to characterize the metabolite profile of a sample fall into two classes: untargeted and targeted metabolomics analyses. Untargeted metabolomics analyses try to measure all of the analytes in a sample, whereas targeted analyses aim to monitor a set of metabolites that are defined and annotated[70]. In this work, I use untargeted metabolomics analyses to better capture the entirety of the metabolic changes within the samples studied, since I do not know which molecules or classes of molecules are important to whole-cell and cell-free bacterial biosensor functionality.

For both targeted and untargeted analyses, advanced analytical equipment is required to measure the metabolites within the samples. The two most popular technologies for this are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) coupled to a separation technique. Due to the vast space of possible chemical structures, no single technology can detect all the metabolites in a sample. As a result, researchers typically focus on a smaller set of molecules that can be detected with a single technique

or combine analytical technologies to cover a larger range of targets. Although these analytical platforms are extensively described elsewhere[71-74], they are briefly described here.

NMR spectroscopy can identify and quantify analytes in a sample without extensive sample workup and is non-destructive[71]. This tool is particularly useful to determine the structural properties of molecules and their concentrations with high reproducibility[72]; however, NMR typically has low sensitivity, which in turn requires large sample sizes. This large sample size requirement can be logistically challenging to meet and limit the utility of NMR for analysis of bacterial biosensor samples.

In contrast, mass spectrometry offers a more sensitive approach that requires smaller amounts of sample[71]. Mass spectrometers consist of an ion source, a mass analyzer, and a detector[73]. The ion source is where the analytes are ionized, which then are sent to the mass analyzer. The mass analyzer separates the ions based on their mass to charge ratio (m/z). Quadrupole, ion trap, and time-of-flight (TOF) mass analyzers are commonly used platforms. Quadrupole mass analyzers separate ions by transmitting them down four parallel cylindrical rods that have been subjected to an electrical field and provide consistent separation[73]. Ion trap mass analyzers separate ions by sending them through to a ring electrode with two end caps that has also been subjected to an electric field. Ion trap analyzers have good sensitivity due to their ability to store ions. TOF mass analyzers separate ions by accelerating them down a flight tube with a constant kinetic energy that allows for mass-based ion separation and has the widest mass range of any mass analyzer. After the ions are sent through one of these mass analyzers, the ions travel to the detector,

which measures the ions, and the resulting output is used to create a spectrum of the relative abundance of the ions as a function of m/z.

Because metabolites are often in complex mixtures, mass spectrometers are commonly coupled to a chromatography-based separation technique to resolve the molecules, allowing for them to be separated by retention time prior to ionization. Liquid and gas chromatography are the most popular separation platforms used. Liquid chromatography-mass spectrometry (LC-MS) separates molecules based on their interactions with mobile and stationary phases and can detect a broad range of metabolites with few required sample preparation steps[74]. LC-MS typically uses soft ionization techniques such as electrospray ionization which cause little to no ion fragmentation; however, this ionization technique can result in adduct ion formations and less informative or comprehensive spectra for analyte identification[75]. Gas chromatography-mass spectrometry (GC-MS) separates molecules based on their boiling points and is ideal for measuring volatile metabolites. To improve nonvolatile metabolite detection, derivatization of samples is typically required using methods such as methoximation and trimethylsilylation, which increase the volatility and thermal stability of metabolites. Although the derivatization process is tedious and time intensive, it is highly beneficial because it broadens the range of metabolites that can be measured with GC-MS. Because GC-MS typically uses electron ionization, which is a hard ionization technique that fragments each ion with high reproducibility, more informative spectra can be obtained[76].

In this work, I use gas chromatography coupled to time-of-flight mass spectrometry (GC-TOF-MS) as the primary analytical tool based on its robust sensitivity and reproducible metabolite spectra. Specifically, I use two-dimensional gas chromatography

coupled to time-of-flight mass spectrometry (GCxGC-TOF-MS), which uses two GC columns with different properties connected to a modulator to more effectively separate molecules that coelute in the first column, improving the instrument's overall resolution and sensitivity[77].

### 1.6 Metabolomics data processing and analysis

Due to the complexity of the data that results from metabolomics measurements, multiple data processing steps and various software packages are required that are often specific to the instrument being used[78]. In general, data processing and analysis entails (1) peak deconvolution, (2) peak alignment and filtering, (3) analyte identification, and (4) statistical analyses. Peak deconvolution aims to separate and create unique spectra for the peaks generated by analytes that elute from the chromatography column at the same time[79]. Peak alignment and filtering work to adjust for instrument drift (which is a common problem that leads to a change in sample baseline position) by aligning the analytes across all the samples and filtering the peaks based on their reproducibility. Analyte identification uses in-house or commercially-available spectral libraries to assign a chemical identity to the chromatographic peaks based on their spectra. Lastly, for statistical analyses, univariate and multivariate analyses can be conducted, and each can contribute significant insights to the interpretation of the data. Univariate analyses focus on one variable at a time and include tests like *t*-tests and analysis of variance (ANOVA). These analyses are useful to assess the individual metabolites in the data set. Multivariate analyses focus on the data set as a whole and include analyses such as principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA)[80]. These analyses are useful to process the

complex data set into a simpler, visual format that can help differentiate between sample classes.

## 1.7 Contribution of this thesis

In this dissertation, I use metabolomics analysis via GCxGC-TOF-MS to address key issues in whole-cell and cell-free bacteria-based biosensor development. For whole-cell biosensors, I explore the metabolic effects of expressing precursor pathways for a pigment reporter system in *E. coli* and identify metabolites that limit this system's utility. I then use this information to move towards addressing some of the limitations by conducting medium supplementation-based validation experiments. For cell-free biosensors, I characterize the endogenous metabolic activity in *E. coli*-based CFE systems, which until my work had never been fully characterized. I compare the metabolic profiles and productivity of differently prepared crude lysates using common preparation methods and characterize the metabolic dynamics within CFE systems. I identify key metabolites that impact CFE reaction activity that are generalizable to differently prepared lysates derived from a variety of strains of *E. coli* and use these insights to improve reaction performance through small molecule and enzyme supplementation efforts, highlighting the importance of understanding the metabolic activity of CFE systems for targeted optimization of cell-free biosensors.

# CHAPTER 2. ANALYSIS OF THE METABOLIC IMPACTS OF AN ENGINEERED LYCOPENE REPORTER SYSTEM ON WHOLE-CELL BACTERIAL BIOSENSORS

Portions of this chapter are reproduced from my publication "Metabolomics analysis of the toxic effects of the production of lycopene and its precursors"[81] in *Frontiers in Microbiology*.

## 2.1 Introduction

The use of cells as microbial factories has significant potential in many different contexts. The intricate enzyme machinery available in nature enables highly specific production of molecules ranging from specialty fuels to pharmaceutical precursors. These approaches also have the potential to be more environmentally friendly and sustainable than synthetic chemistry and petrochemical approaches, in aspects ranging from solvent to energy usage. The field of synthetic biology continues to provide tools that advance the forefront of what chemicals cells are capable of synthesizing and how much they are capable of making, with ever-increasing titers of ever-more-complex molecules.

Metabolomics, the systems-scale study of the biochemical intermediates of metabolism, can help to inform the development of engineered strains[82] through characterization of key endpoints and small molecule regulators of cellular state[83, 84]. While metabolomics has proven to be useful in complementing the development of strains designed to maximize biochemical production[85-87], it has been largely unexplored for

biosensor development, which instead focuses on selectively producing high yields of reporter molecules in response to a target analyte.

In recent work, we have sought to develop low-cost, minimal-equipment biosensors for application in the developing world[88]. Specifically, we have used *E. coli* as the chassis organism for a whole-cell biosensor to measure zinc status in blood samples, with biosynthesis of differently colored pigments used as the readout for the sensor. In this application, the key challenges to date have included the precise control of production of only one pigment at a time, and the complete repression of any pigment production to enable a dense, colorless initial inoculum that allows a rapid output response[89]. Our lab has pursued multiple strategies to limit unwanted pigment production and speed up visible pigment production once desired, including precursor supplementation.

Specifically, we engineered strains that produced negligible visible lycopene in the uninduced state and produced unmistakably visible lycopene within approximately three hours after induction. To attempt to speed up this response time, we heterologously expressed in the same strain the mevalonate pathway, to overproduce the precursors for lycopene biosynthesis. We expected that by providing a mechanism for increased production of lycopene precursors, we could increase total titers and/or the time necessary for unmistakably visible lycopene production. Furthermore, we expected that by inducing the production of these colorless precursors overnight, the cell would be primed for lycopene production upon induction of the heterologous lycopene biosynthesis pathway.

To our surprise, the effect of mevalonate pathway supplementation on cell growth and specific lycopene production varied greatly depending on the time of supplementation.

Specifically, if the mevalonate pathway was induced overnight before inoculation and induction of lycopene biosynthesis, specific lycopene production increased, and cell growth decreased significantly. In contrast, if the mevalonate pathway was only induced at the same time as the lycopene biosynthesis pathway, there was minimal impact on cell growth and an even greater increase in specific lycopene production. While toxicity associated with lycopene production has certainly been reported before[90-92], in this case our results suggested that it was not necessarily the total levels of lycopene that were causing growth inhibition. To enable further engineering of this pathway to speed up lycopene production, it was of great interest to characterize the metabolic phenotypes underlying this curious cellular behavior.

In this chapter, we use metabolomics (via two-dimensional gas chromatography coupled to mass spectrometry, GCxGC-MS) to characterize the underlying metabolic states across these counterintuitive observations. We focus on the central carbon metabolites with broad functional and metabolic impacts that are well-measured by GCxGC-MS. After briefly presenting an improved method for quenching *E. coli* for metabolomics analysis, we first characterize the metabolic state of our baseline lycopene-producing strain in the growth conditions characteristic of our biosensor application. We then study the induction of the mevalonate pathway at inoculation and during the overnight culture before induction, with and without the lycopene biosynthesis genes in the same strain. We briefly consider some of the most noteworthy metabolic changes in the affected strains and conditions, suggesting potential hypotheses for the regulation or mechanisms mediating the observed behaviors, and then test and validate one of those hypotheses via a medium supplementation experiment.

## 2.2 Experimental Methods

### *2.2.1 Strains characterized*

*Escherichia coli* K-12 DH10B (New England Biolabs, Ipswich, MA) was used as the host strain for all metabolite production experiments, and all heterologous proteins were expressed from standard expression plasmids. A detailed description of plasmid assembly can be found in McNerney, et al[89]. In short, the lycopene producing genes *crtE*, *crtB*, and *crtI* were amplified from Part bba_k274100 of the Registry of Standard Biological Parts and placed under the control of an IPTG-inducible promoter. The mevalonate pathway genes were amplified from the plasmid pJBEI-6409[93], which was obtained from Addgene (Cambridge, MA, USA) and placed under control of the arabinose-inducible promoter pBAD to create the plasmid *pBadMEV*. The high lycopene producing plasmid *pLac32EBI* has the medium-strength RBS B0032 on the *crtE*, *crtB*, and *crtI* genes, and the weak lycopene producing strain *pLac33EBI* has the weak RBS B0033 on the *crtE*, *crtB*, and *crtI* genes. In all experiments comparing the effect of mevalonate pathway supplementation, the high lycopene-producing plasmid was used. Plasmids *pLacØ* and *pBadØ* were constructed as controls and contain no coding sequence after the promoter.

### *2.2.2 Cell culture*

In experiments comparing the metabolic changes based on lycopene production levels, cells were transformed with individual plasmids. In experiments comparing the effect of the mevalonate pathway, cells were cotransformed with two plasmids. Following transformation, cells were plated on LB plates with appropriate antibiotics for selection and grown at 37 °C overnight. Freshly transformed colonies were then inoculated in

triplicate into LB medium with the appropriate inducers and grown at 37 °C and 180 rpm for 18 hours. Cultures were then concentrated and inoculated to an OD of 0.2 in fresh medium, the appropriate inducers were added, and the sample was aliquoted into culture tubes corresponding to different time points. Samples were analyzed at the time of inoculation and at hours 1, 2, 4, and 6. At each time point, optical density and lycopene content were measured, and samples were collected for metabolomics analysis. Optical density was quantified by measuring the absorbance at 600 nm in a ThermoFisher Genesys 20 spectrophotometer with a 10 mm path length.

An additional triplicate set of *pLac32EBI+pBadØ* samples as well as two triplicate sets of wild type were grown during the lycopene production comparison experiment to control for variation between experiment days. Lycopene extractions and quenching of these samples were performed with those from the lycopene production comparison experiment, but one set of wild-type samples as well as the additional set of *pLac32EBI+pBadØ* samples were stored in a -80°C freezer until extraction of the second experiment's samples.

LB medium composed of 10 g $L^{-1}$ NaCl, 5 g $L^{-1}$ yeast extract, and 10 g $L^{-1}$ tryptone was used in all experiments. Either 1 mM IPTG or 0.01% (w/v) arabinose was used for induction, and the following antibiotics were used for appropriate selection: tetracycline (15 μg/mL), kanamycin (30 μg/mL), and carbenicillin (100 μg/mL).

### *2.2.3 Lycopene extraction and HPLC analysis*

Lycopene was extracted from cultures and analyzed as described previously[89]. Briefly, 500 μL of bacterial culture was pelleted and resuspended in 50 μL of ultrapure

water. Lycopene was extracted with 1 mL of acetone at 50 °C for 20 minutes. Cellular debris was pelleted, and the supernatant was removed for analysis. Sudan I (TCI America, Portland, OR) was used as an internal standard[94] and added to the acetone used for extractions at a concentration of 1 μg/mL.

All HPLC analysis was conducted on a Shimadzu Prominence UFLC using an Agilent C18 4.6 mm x 50 mm column with a 5 μm particle size and a Shimadzu photodiode array detector. A solvent ratio of 50:30:20 acetonitrile:methanol:isopropanol was used as the mobile phase[95] and run at a flow rate of 1 mL/min with a 25 μL sample injection volume. Absorption was detected at 471 nm. Retention times and peak intensities were compared to an analytical lycopene standard (Millipore Sigma, St. Louis, MO) spiked into control extractions from DH10B cells, and the internal standard Sudan I was used to account for acetone evaporation during the extraction protocol and for instrument drift.

### *2.2.4 Quenching*

This method is a modified version of the methods in Spura *et al* and Yasid *et al*[96, 97]. A quenching solution composed of 30% ethanol (v/v) and 0.6% NaCl (w/v) in a 15 ml or 50 mL conical was cooled to -15 °C. Each tube was prefilled with a certain amount of quenching solution for a 2:1 quench:sample ratio based on estimated OD at sampling time.

Samples were taken at 0 hours, 1 hour, 2 hours, 4 hours, and 6 hours for both experiments to capture the metabolic profile over a relevant timeframe for lycopene production. The sample was added to each tube and quickly mixed by inversion. Each tube was kept in a -15 °C 70% methanol bath until the sample temperature reached -5 to -8 °C. The tubes were then centrifuged for 5 minutes at 3500 rpm at -10 °C and transferred back

to a -15 °C bath. The supernatant was removed, and the tubes were frozen in liquid nitrogen. The samples were stored in a -80 °C freezer until extraction.

### *2.2.5 Metabolite extraction*

The extraction method is a modified version of the freeze-thaw method in Faijes *et al* and Yasid *et al*[97, 98]. The cell pellet was resuspended in 500 µL of -80 °C methanol and transferred to a 1.5 mL microcentrifuge tube. The suspension was then frozen in liquid nitrogen, thawed on ice, and centrifuged for 2 minutes at 10,000 g at 4 °C. The supernatant was collected and stored in a separate microcentrifuge tube. The pellet was again suspended in an additional 500 µL of -80 °C methanol, frozen, thawed, and centrifuged under the same conditions as the previous step. The supernatant was collected, and the pellet was resuspended with 250 µL of cold water, undergoing this freeze-thaw process for a final time. Once the supernatant was collected, the pooled supernatants were more accurately normalized to OD and were transferred to a CentriVap to be centrifugally concentrated at 40 °C until completely dry. The dried samples were stored in a -80 °C freezer for later processing.

### *2.2.6 GC-MS analysis*

Before derivatization, the samples were transferred to a CentriVap to be dried at 40 °C for 15 min. Samples were derivatized as previously described[99, 100]. 10 µL of 20 mg/mL *O*-methylhydroxylamine hydrochloride (MP Biomedicals, LLC, Santa Ana, CA) in pyridine was added to each dried sample and shaken at 1400 rpm for 90 minutes at 30 °C. 90 µL of *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA) + 1% trimethylchlorosilane (TMCS) (Thermo Scientific, Lafayette, CO) was then added to the

samples and shaken at 1400 rpm for 30 minutes at 37 °C. Samples were centrifuged at 21,100 *g* for 3 minutes, and 50 µL of the supernatant was added to an autosampler vial. Samples were spiked with 0.25 µL of a retention time standard solution composed of fatty acid methyl esters (FAMES) and an internal standard of nonadecanoic acid methyl ester dissolved in dimethylformamide. In parallel, a quality control (QC) sample was prepared by removing 150 µL of extract from each sample and aliquoting 1.15 mL for experiments comparing the metabolic changes based on lycopene production levels. For experiments comparing the effect of the mevalonate pathway, 75 µL was removed. 1.15 mL and 0.65 mL of QC was aliquoted for each experiment respectively, dried, and derivatized with each batch of 9-10 samples. At the beginning of the GC-MS run, the QCs were injected once and repeated again after every 4 to 5 sample injections to allow for downstream correction for batch effects. A derivatization blank was prepared and run with every batch of samples.

A LECO Pegasus 4D instrument with an Agilent 7683B autosampler, Agilent 7890A gas chromatograph and time-of-flight mass spectrometer (TOF-MS) was used to analyze the samples. The first column was an HP-5, 28 m long × 0.320 mm ID × 0.25 µm film thickness (Agilent, Santa Clara, CA, USA), and the second was an Rtx-200, 1.75 m long × 0.25 mm ID × 0.25 µm film thickness (Restek, Bellefonte, PA, USA). More detailed gas chromatography, autosampler, and mass spectrometry methods are provided in Appendix A.

### *2.2.7 Data analysis*

Sample runs were analyzed in ChromaTOF (LECO, St. Joseph, MI) to determine baseline, peak area, and peak identification as described previously[101, 102]. Briefly, settings included a baseline offset of 0.5, automatic smoothing, 1st dimension peak width of

36 seconds, 2nd dimension peak width of 0.10 seconds, and a match of 700 required to combine peaks with a minimum signal-to-noise (S/N) of 5 for all subpeaks. Peaks were required to have a S/N of 10 and have a minimum similarity score of 800 before assigning a name. Unique mass was used for area and height calculation.

MetPP (http://metaopen.sourceforge.net/metpp.html) was used to align the samples[103]. Sample files and a derivatization reagent blank file were uploaded from ChromaTOF. Unknowns were retained during the peak alignment process. The derivatization reagent blank file was used to subtract peaks resulting from the sample preparation reagents from the corresponding cells' sample files. On-the-fly alignment was used with manually selected quality control samples as the peak list for primary alignment. Due to the size of the mevalonate induction time variation experiment, MetPP could not handle on-the-fly alignment using all of the quality control samples. Instead, one quality control sample was used from each batch as well as additional quality controls from the first, last, and fifth batches to perform on-the-fly alignment. Peak alignment was performed using the default criteria.

After alignment, further processing of the data was done based on the procedure previously described[104]. Batch effects were removed from the data set using LOESS for both experiments. To remove analytes that were not reproducibly detected, analytes for which more than half of the values were missing in the QC samples or for which the QC samples had a coefficient of variance larger than 0.5 were removed from the data set. Then, missing values were manually corrected using small value correction only if all the values were missing in the biological replicate.

Samples from the mevalonate induction time experiment were processed with Combat[105] to remove batch effects that were evident from principal component analysis of the initial data.

Finally, MetaboAnalyst (http://metaboanalyst.ca/) was used for statistical and pathway analysis[106]. For both analyses, remaining missing values were k-nearest neighbors (KNN) corrected. Data was filtered using the interquantile range method and then log-transformed using generalized logarithm transformation (base 2) and autoscaled. Differences were considered significant at false discovery rate-corrected $p < 0.05$. The metabolomics datasets for this study have been deposited to Metabolights with the dataset identifier MTBLS642.

All samples from the experiment comparing the metabolic effects of the amount of lycopene production, except for one triplicate wild type set and the pLac32EBI+pBadØ, were extracted, derivatized, and analyzed together. All samples for the experiment comparing the metabolic effect of the mevalonate pathway as well as those excluded from the previous experiment were extracted, derivatized, and analyzed together.

### *2.2.8 Methionine supplementation*

A triplicate set of the strains *pLac32EBI+pBadMEV* and *pLac32EBI+pBadØ* were cultured and induced as described above with the exception of the addition of 2mM methionine to the medium at the start of the experiment. A triplicate control set of the same strains under the same culture and induction conditions were grown without methionine to test for differences in growth due to methionine addition. Each sample's optical density

was measured at the time of inoculation and at hours 1, 2, 4, and 6, and was quantified as described above.

## 2.3 Results

### *2.3.1 Modified quenching protocol*

Fast quenching of cellular metabolism for metabolomics analysis is critical for acquisition of samples that accurately represent cellular state in culture rather than artifacts induced by sample processing. Fast filtration, cold methanol, and cold ethanol are the three most common quenching methods for *E. coli*. However, due to the time needed for filtration and quenching of each sample, fast filtration is not appropriate for measuring metabolites with high turnover rates[107]; it has also been reported to cause leakage in Gram-negative bacteria[108]. Numerous studies have shown that the cold methanol method also causes serious leakage from the cell[108-111]. We thus selected the cold ethanol method by Spura *et al*[96], which causes significantly less leakage.

Following the original protocol, combining a 37 °C sample with a -20 °C quenching solution at a 1:1 ratio lowers the sample's temperature only to about 7-10 °C instantaneously before further cooling to -5 to -10 °C in a cold bath. 7-10 °C is clearly insufficient to halt metabolism instantaneously, and the longer the cells spend in quenching solution, the more leakage is likely to occur[112]. To lower this instantaneous post-quenching temperature and reduce the time the cells need to stay in the solution, Yasid *et al* modified Spura's method by decreasing the quenching solution's temperature to -35 °C[97]. While this resolves the instantaneous quenching temperature issue (yielding temperatures of 1 to -2 °C upon quenching), the quench solution is an icy slurry that melts in a non-uniform

fashion and thus requires extra manual supervision that would not be feasible at the scale of sampling needed for this experiment.

To avoid these issues, we modified the quenching solution to allow for a rapid sample temperature decrease to about 0 °C while avoiding ice slurries and manual supervision of each quenched sample. We increased the ratio of quenching solution volume to sample volume while maintaining the original overall percentage of ethanol and salt in the final quenched sample (40% ethanol/0.8% sodium chloride quenching solution added at a 1:1 ratio to yield a 20% ethanol/0.4% sodium chloride quenched sample). Our modified quenching solution consists of a 30% ethanol (v/v)/0.6% sodium chloride (w/v) solution added at a 2:1 ratio to sample. Although decreasing the amount of ethanol in the solution increases the freezing point to about -20 °C, we were able to keep our quenching solution at -15 °C without any ice formation while lowering the sample's temperature upon initial mixing to about 1 to -3 °C (Table 1).

**Table 1: Quenching solution comparisons.**

| | Quenching solution temperature (°C) | Sample temperature (°C) | Temperature after mixing (°C) | STD (°C) |
|---|---|---|---|---|
| Spura *et al* Method | -20 | 37 | 8.4 | 1.1 |
| Yasid *et al* Method | -35 | 37 | -0.4 | 1.4 |
| Modified Method | -15 | 37 | -1.1 | 1.9 |

### *2.3.2 Lycopene levels of selected strains have minimal impacts on growth characteristics*

Overcoming the toxicity effects of heterologous expression of the lycopene biosynthesis pathway from *Pantoea ananatis* is a challenge that has been encountered

repeatedly in the field of metabolic engineering. In our previous work, we developed *E. coli* strains that could express different levels of lycopene with minimal effects on cell growth by replacing the native ribosomal binding sites (RBSs) on the *crtEBI* genes with RBSs of varying, but weaker, strength[89].

Nonetheless, it has been previously reported that systems-scale metabolic changes due to lycopene production can be detected even before lycopene levels are measurable[113]. While it is obvious that metabolites in the lycopene biosynthesis pathway would have different profiles in strains with different lycopene production potential, we were particularly interested in measuring the extent to which metabolic changes were detectable in more central portions of metabolism with broader impact on cell function and phenotype. To this end, we sought to perform GCxGC-MS metabolomics on two different strains, *pLac32EBI* and *pLac33EBI*, with higher and lower (respectively) levels of lycopene production, and a non-lycopene-producing wild type strain.

All three strains were grown and sampled over 6 hours after inoculation and induction from a saturated overnight culture, as this time-frame and protocol are relevant to the biosensor application for which the cells are producing lycopene. The growth profiles of the strains indicate that lycopene production at levels considered in this work has a measurable but quantitatively small impact on growth under these culture conditions when compared to wild type (Figure 4A). This is consistent with our observations from previous work in this area. Differences in growth curves between the higher and lower-producing strains were generally insignificant. Despite the nearly identical growth profiles of the two lycopene-producing strains, their total lycopene levels differ by approximately an order of magnitude throughout most of the time course (Figure 4B & Table 2), suggesting that while

lycopene production can be toxic to *E. coli* growth at sufficiently high levels, the cells under study here have quite a bit of tolerance for it. Confirming this fact then allowed us to attribute subsequent differences in metabolite profiles between strains to differences in lycopene production rather than non-specific metabolic responses to toxicity.

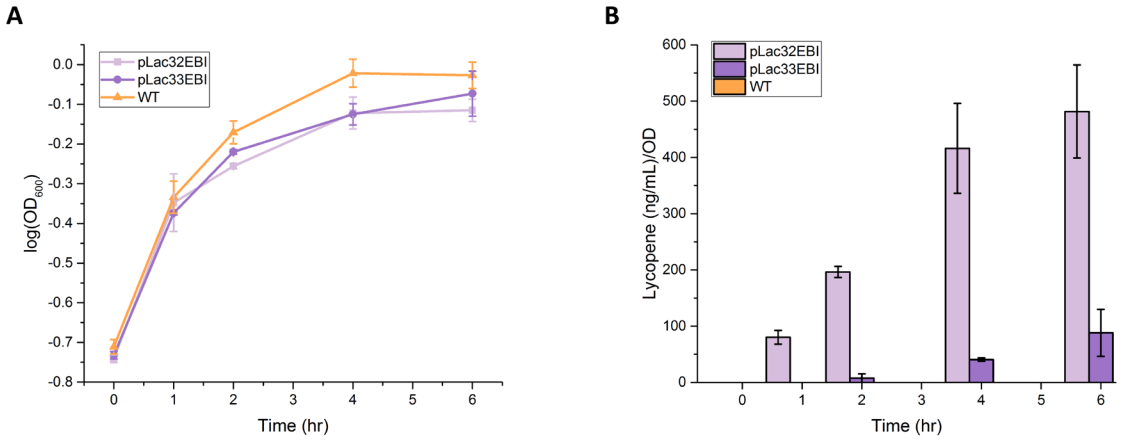


**Figure 4: Growth curve and lycopene production of *pLac32EBI*, *pLac33EBI*, and wild type.**
**(A) Although the two lycopene-producing strains have nearly identical growth profiles, the production of lycopene does have a minor impact on growth. (B) OD-normalized lycopene production is different by about an order of magnitude over most of the time course. Error bars represent standard deviation.**

**Table 2: OD-normalized lycopene production for the high and low lycopene producing strains *pLac32EBI* and *pLac33EBI* at each time point.**

| Strain | Lycopene (ng/mL)/OD | | | | | Lycopene Production Rate ((ng/mL)/OD)/ hr |
|---|---|---|---|---|---|---|
| | 0 hr | 1 hr | 2 hr | 4 hr | 6 hr | |
| *pLac32EBI* | 0 | 80.3 ± 12.2 | 196.3 ± 9.8 | 416.5 ± 79.9 | 481.5 ± 82.6 | 85.1 |
| *pLac33EBI* | 0 | 0 | 7.6 ± 7.6 | 40.6 ± 3.0 | 88.2 ± 41.7 | 15.2 |

### *2.3.3 Culture time and variation in lycopene production levels affect metabolism*

To test whether this degree of variation in the production of lycopene has measurable effects on metabolism, we performed metabolomics analysis on all three strains (two lycopene producers and the wild type). Two-dimensional gas chromatography coupled to mass spectrometry (GCxGC-MS) was used to analyze the metabolic profiles of the intracellular samples collected over the six-hour time-course experiments; this technique is particularly effective for the analysis of small polar metabolites such as those in central carbon metabolism. After peak alignment, data processing, and removal of peaks that were not reproducibly measured, all 1002 remaining peaks – which included both identified and unannotated analytes – were used in subsequent analyses.

We found that the predominant effect on the metabolic profiles of each strain was the time course rather than the strain. We used principal component analysis (PCA) as an unsupervised dimensional reduction approach to characterize metabolic changes across strains and time points and identify the dominant axes of variation in the dataset, along with the metabolites that most strongly contribute to those axes of variation. Directly comparing the higher and lower producers (Figure 5A) and the higher producer to wild type (Figure 5B), it is clear that the two major modes of variation (referred to as principal components (PCs)), accounting for over 40% of the variability in the data, are affected more by the time in culture than by the lycopene expression levels of the strains. There is little to no separation between strains across the time course. Similar analysis of all three strains together yields the same results (Figure 6). (The initial inoculation time point was excluded from these plots for clarity since it overlaps in PCA space for all samples (Figure 7)). Consistent with this, two-way analysis of variance (ANOVA2) on all strains for hours

1 to 6 yielded hundreds of measured analytes with a significant time effect but fewer than 100 with significant strain effects (Table 3). While metabolic changes are dominated by time-specific changes, there are still detectable and measurable differences between the three strains metabolically.

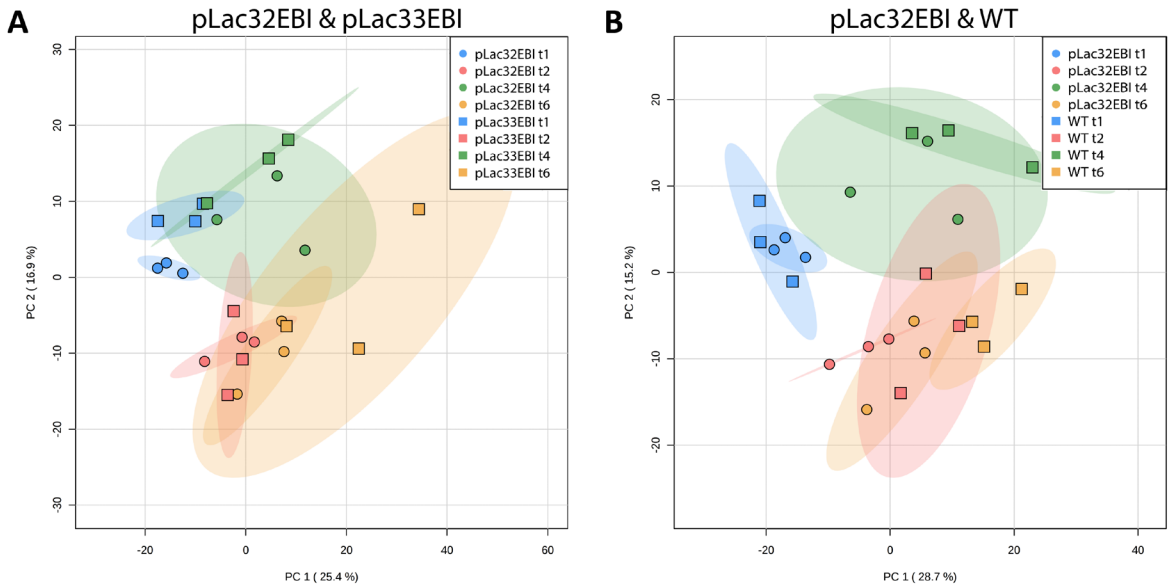


**Figure 5: PCA of high and low lycopene producing strains.**
**(A) pLac32EBI and pLac33EBI and (B) pLac32EBI and wild type metabolomics data were compared. (A) PCA shows little to no separation between the pLac32EBI and pLac33EBI strains across time points, indicating that variance is affected more by time than the amount of lycopene production. (B) There is also little to no separation between the pLac32EBI and wild type strains across the time course, indicating that variance is affected more by time than the production of lycopene. All colored ellipses represent 95% confidence intervals for each group.**

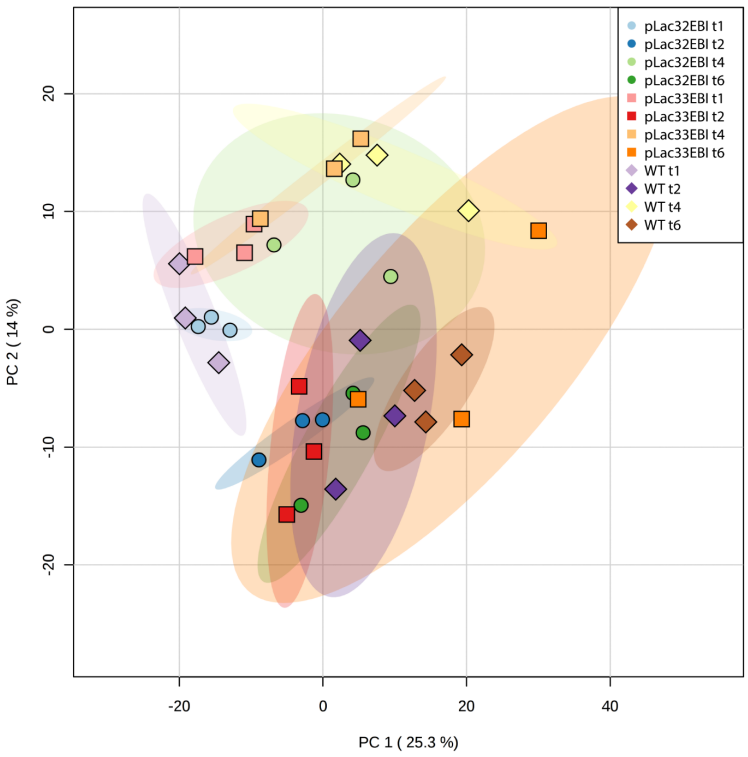


**Figure 6: PCA of *pLac32EBI*, *pLac33EBI*, and wild-type metabolomics data.**
**PCA of all three strains indicate that there is little to no separation between strains across the time course. All colored ellipses represented 95% confidence intervals for each group.**

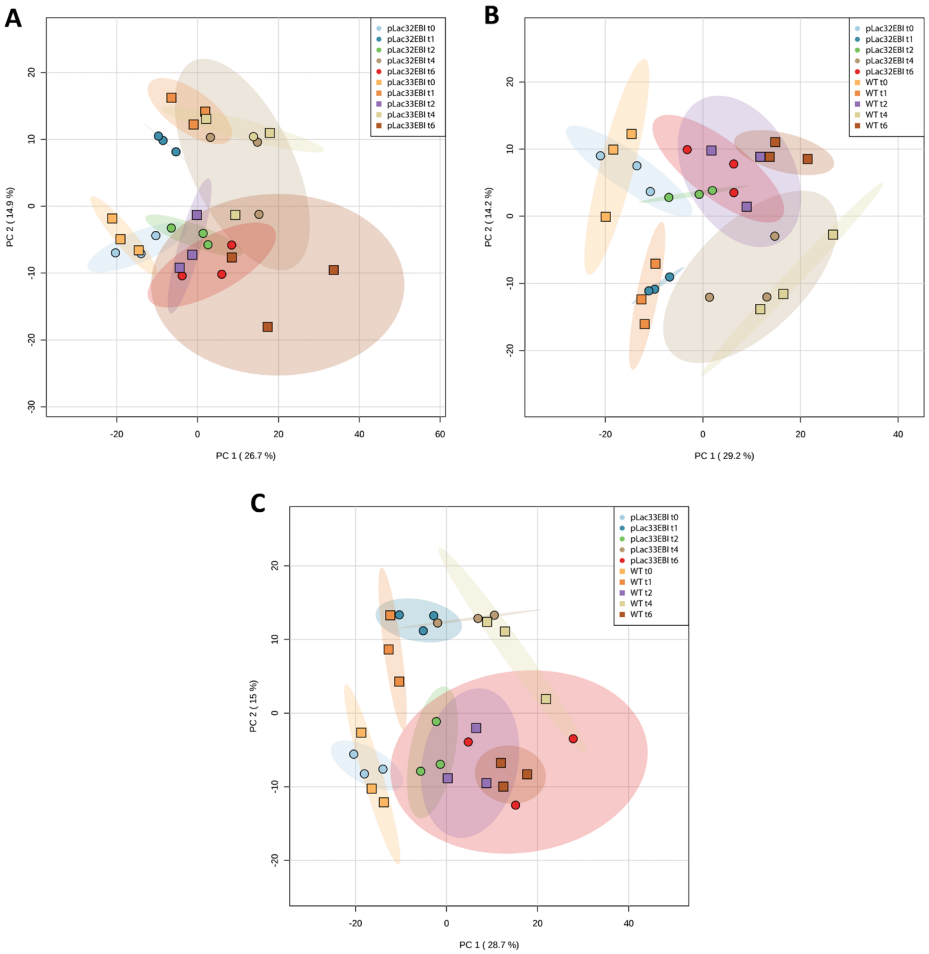


**Figure 7: PCA of high and low lycopene producing strains at all time points.** **(A) *pLac32EBI* and *pLac33EBI,* (B) *pLac32EBI* and wild-type, and (C) *pLac33EBI* and wild type metabolomics data were compared at all time points. In all analyses, strain groups at the initial inoculation time point do not separate from one another. Based on this, the initial inoculation time point was excluded in subsequent PCA analyses and from figures in the main text. All colored ellipses represent 95% confidence intervals for each group.**

**Table 3: Two-way ANOVA for *pLac32EBI* and *pLac33EBI* as well as *pLac32EBI* and wild type from 1-6 hours.**

| Strains | Group Effects | Time Effects | Group & Time Effects | Interaction Effects |
|---|---|---|---|---|
| *pLac32EBI* and *pLac33EBI* | 84 | 290 | 7 | 0 |
| *pLac32EBI* and wild type | 50 | 360 | 29 | 0 |

Figure 8 shows PCA plots at individual time points, during which the three strains are easily distinguishable via their metabolic profiles at 6 hours, and distinguishable to a modest extent at earlier time points. Taken together, these results indicate that metabolic changes associated with varying levels of lycopene production are, while detectable, less quantitatively significant than the changes induced as a function of time during batch growth.

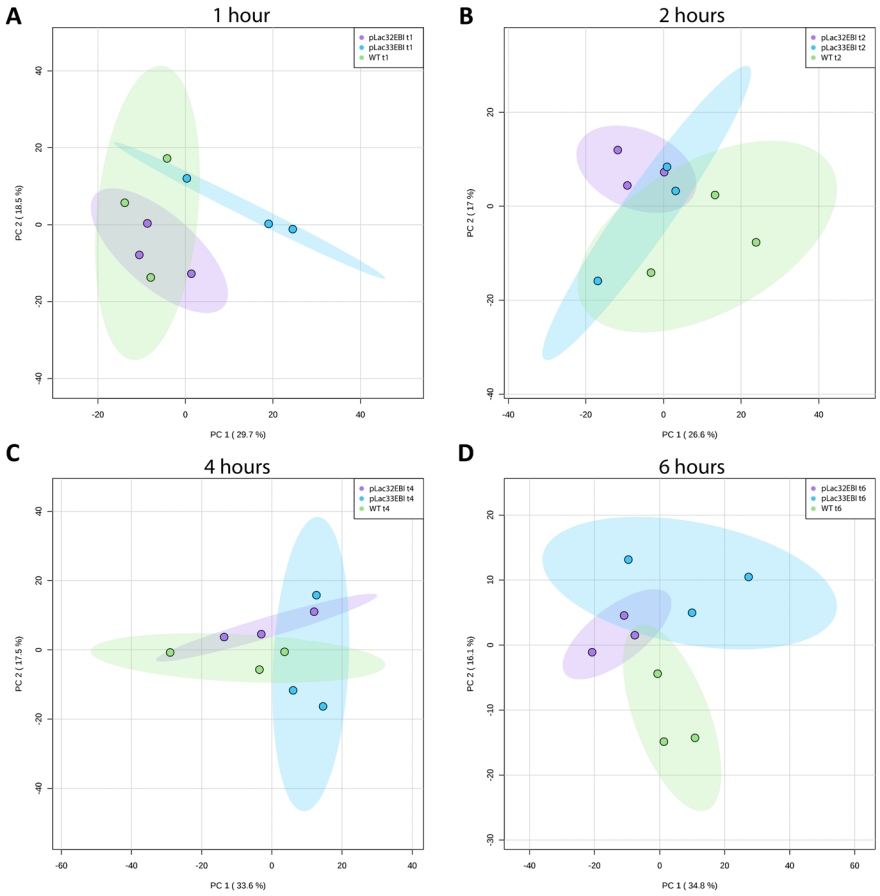


**Figure 8: PCA of high and low lycopene producing strains at 1, 2, 4, and 6 hours. *pLac32EBI, pLac33EBI*, and wild type metabolomics data were compared at each time-point. At (A) 1 h, (B) 2 h, and (C) 4 h, there is a small degree of separation between strains. At (D) hour 6, there is clear separation between all strains. All colored ellipses represent 95% confidence intervals for each group.**

We then analyzed through one-way ANOVA at each individual time point which metabolites were significantly changed across the strains. Of the individual ANOVA-significant analytes across the time points, most were unannotated, with only a handful of putatively annotated metabolites. These known metabolites did not appear to have any strong relationship to lycopene biosynthesis or other strong pathway enrichment, precluding a more detailed or mechanistic interpretation of the significantly changing metabolites. In addition, previously-reported metabolic indicators of cell stress response in *E. coli* (for example, increased production of almost all amino acids and decreases in L-alanine and L-methionine[114-116]), were generally not evident in our data (with a few minor exceptions shown in Figure 9).

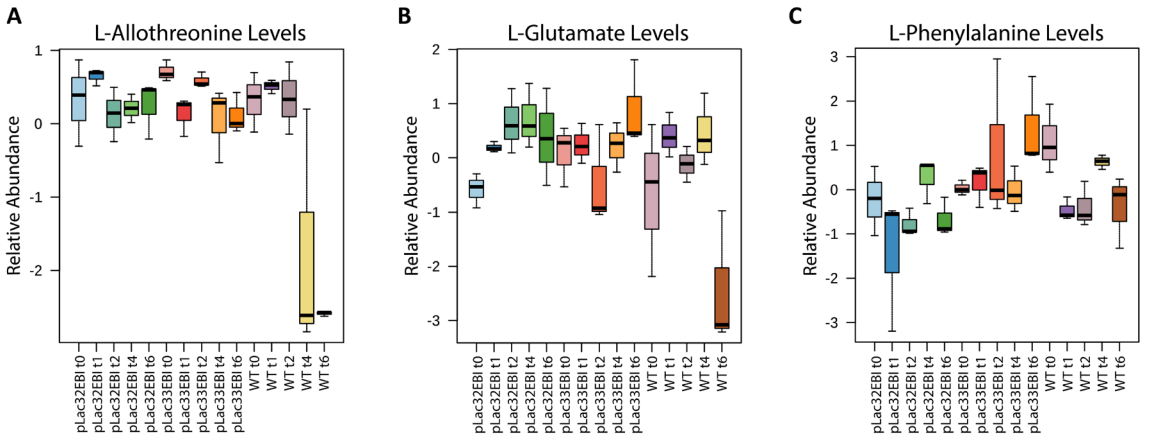


**Figure 9: Profiles of some amino acids of potential interest.**
**(A) L-allothreonine decreased in wild type but not in the lycopene-producing strains. (B) This trend was also seen for L-glutamate levels. (C) At hour 6, *pLac33EBI* had higher L-phenylalanine levels than the other two strains. Box and whisker plots depict the normalized peak areas. Black lines are the medians, and boxes are the middle 50% values. Error bars represent standard deviation.**

### *2.3.4 Time of mevalonate pathway induction significantly affects growth and lycopene production*

While the baseline characterization of the metabolic impacts of lycopene production on the cells was important, our main goal was to study the cellular response to induction of the mevalonate pathway. The mevalonate pathway supplements the production of FPP, the last endogenous precursor to lycopene in *E. coli*. Our previous work showed that inducing the mevalonate pathway in overnight pre-culture drastically decreases both cell growth and per-cell lycopene production compared to induction at inoculation of the 6-hour time course[89].

To study this phenomenon, we used just the higher-producing strain (*pLac32EBI*) as a model because it provided easily measurable levels of lycopene relevant for our target application but avoided significant growth toxicity. We then heterologously expressed either the lycopene production pathway, the mevalonate pathway, or both in three strains (*pLac32EBI+pBadØ*, *pLac32Ø+pBadMEV*, and *pLac32EBI+pBadMEV*, respectively). In all cases, the lycopene-producing pathway (or its null construct) was induced only at inoculation after overnight culture, while the mevalonate pathway (or its null construct) was induced either only at inoculation or starting at the overnight pre-culture.

Consistent with our previous study, overnight induction of the mevalonate pathway in the lycopene-producing strain severely inhibited growth compared to induction at inoculation for the identical strain (Figure 10A). Even for the overnight-induced mevalonate-only strain, there was visible growth inhibition compared to induction at inoculation for the identical strain. Of note is that this growth inhibition seemed to subside

between 2 and 4 hours into the culture, with the cells returning to a growth rate similar to that of any other non-growth-restricted condition. Thus, overnight induction of the mevalonate pathway has some intrinsically toxic effect on cell growth that can be overcome after a sufficient recovery time, but production of lycopene prevents any such recovery and exacerbates the toxicity.

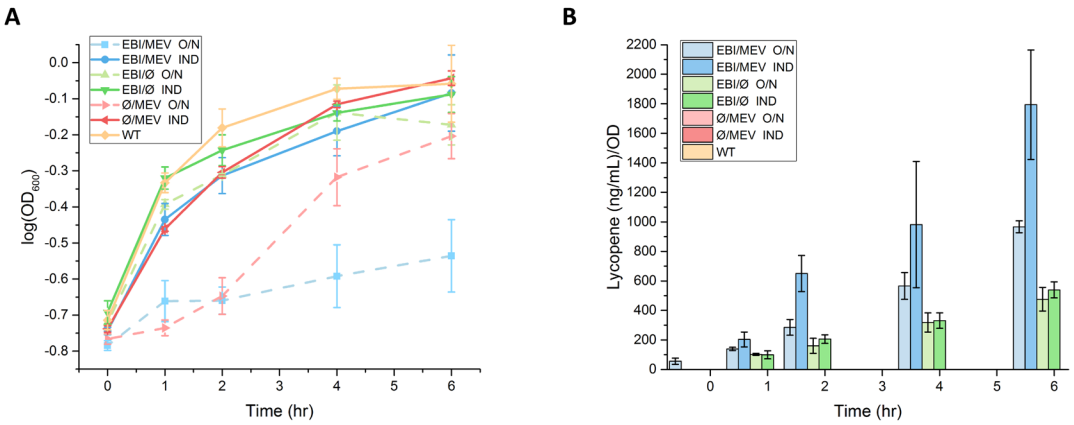


**Figure 10: Growth profiles and lycopene production in response to overnight and inoculation induction of the mevalonate pathway.**
**(A) When the mevalonate pathway is induced during overnight growth (O/N, dashed lines), mevalonate-expressing strains suffer from growth inhibition. The growth inhibition subsides at the end of the time course for the mevalonate-only strain, but persists for the strain also expressing lycopene. All strains have little to no growth inhibition when mevalonate is induced at inoculation (IND, solid lines). (B) OD-normalized lycopene production is greater for the mevalonate-expressing strains regardless of induction time, but overnight induction yields lower OD-normalized lycopene production than inoculation induction. Error bars represent standard deviation.**

Nonetheless, this toxicity is not directly attributable to lycopene levels in the cells. Consistent with previous observations, overnight induction of the mevalonate pathway increased normalized lycopene production and decreased growth rate, while mevalonate pathway induction at the same time as lycopene pathway induction yielded even higher normalized lycopene levels and less cell toxicity (Figure 10B & Table 4). Thus, potential

toxicity of lycopene or its intermediates appears to be insufficient to explain the increased growth restriction due to mevalonate pathway induction.

**Table 4. OD-normalized lycopene production and rate for lycopene producing strains *pLac32EBI+pBadMEV* and *pLac32EBI+pBadØ* at each time point.**

| Strain | Lycopene (ng/mL)/OD | | | | | Lycopene Production Rate ((ng/mL)/OD) /hr |
|---|---|---|---|---|---|---|
| | 0 hr | 1 hr | 2 hr | 4 hr | 6 hr | |
| *EBI/MEV* (O/N) | 55.7 ± 21.5 | 138.8 ± 10.9 | 284.9 ± 52.9 | 566.0 ± 90.6 | 967.0 ± 40.3 | 152.7 |
| *EBI/MEV* (IND) | 0 | 203.2 ± 50.2 | 650.0 ± 122.3 | 981.4 ± 427.9 | 1794.4 ± 370.9 | 291.4 |
| *EBI/Ø (O/N)* | 0 | 101.9 ± 6.9 | 159.8 ± 51.2 | 318.0 ± 64.6 | 475.6 ± 79.8 | 77.7 |
| *EBI/Ø (IND)* | 0 | 99.3 ± 27.0 | 205.7 ± 29.5 | 330.8 ± 52.0 | 539.0 ± 54.2 | 86.8 |

### *2.3.5 Time of mevalonate pathway induction significantly affects metabolism*

To characterize the underlying metabolic impact of the different heterologously expressed pathways and induction times, we performed metabolomics analysis, as described above, on the engineered strains in the two induction conditions. After peak alignment, data processing, and removal of peaks that were not reproducibly measured, all 400 remaining peaks – which included both identified and unannotated analytes – were used in subsequent analyses. PCA for samples from 1 hour through 6 hours clearly demonstrate the metabolic impact that the mevalonate pathway has on the cells. (Similar to the previous experiment, the initial inoculation time point was omitted from these plots

for clarity since it overlaps in PCA space for all samples (Figure 11). This similarity of initial metabolic profiles is perhaps unsurprising, as the steps leading up to this sampling, including saturation in stationary phase in depleted growth medium followed by centrifugation and resuspension during an extended time at room temperature, may have induced a common metabolic response across all of the conditions.) The *pLac32EBI+pBadMEV* strain shows essentially complete separation in the first principal component strictly by induction time (Figure 12A). This difference is likely not attributable just to differences in lycopene production levels, as similar separation was not seen between the higher- and lower-producing strains in Figure 5A. The *pLac32Ø+pBadMEV* strain also exhibits induction time-dependent group separation (Figure 12B), though not to the same extent; the two treatment conditions do not completely separate out in PCA space, but for any individual time point the two treatment conditions are obviously separated in the first principal component. The *pLac32EBI+pBadØ* strain, on the other hand, showed no visible separation between groups at any time point (Figure 12C).

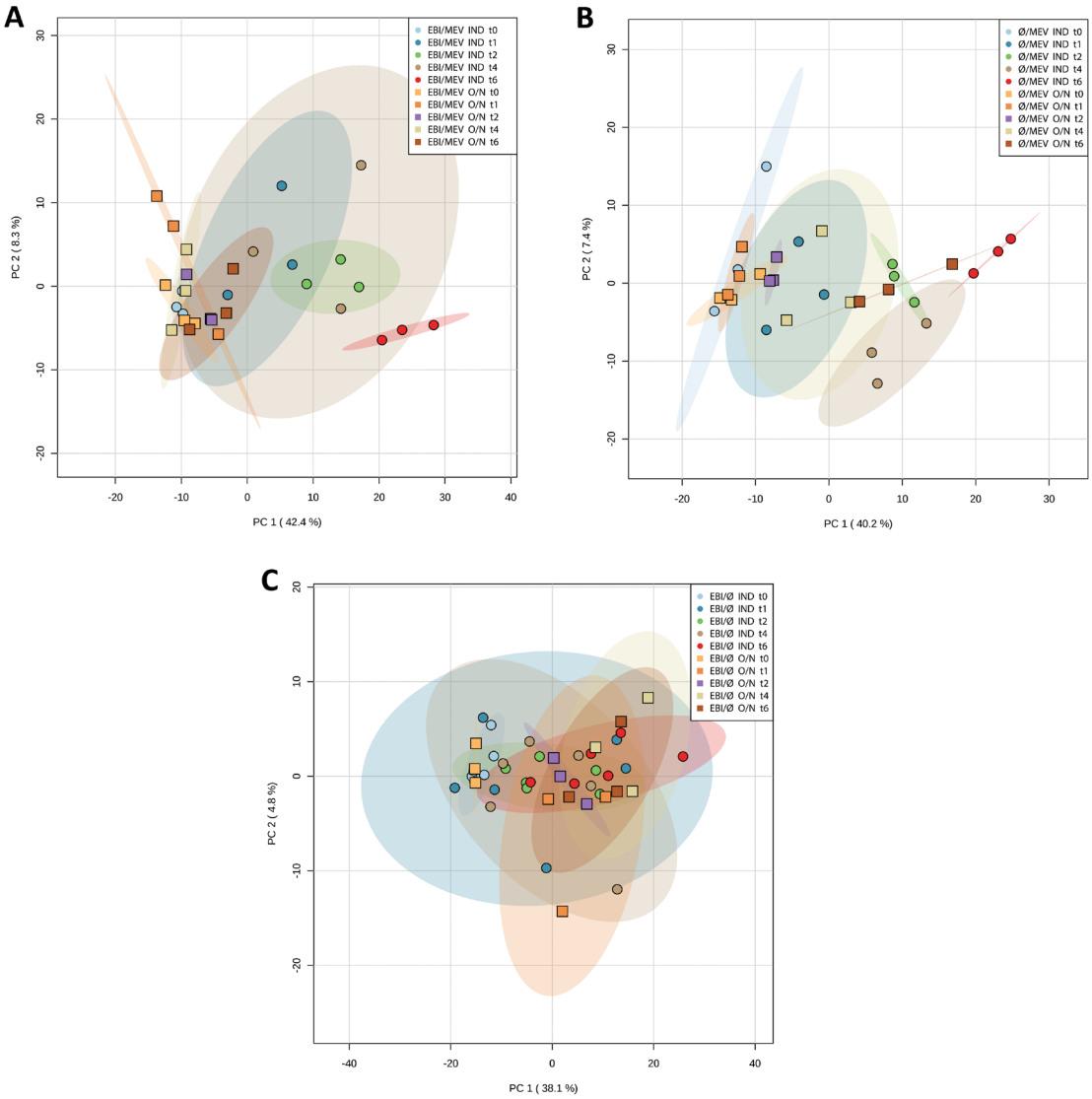


**Figure 11: PCA of overnight- and inoculation-induced strains.**
**(A) pLac32EBI+pBadMEV, (B) pLac32Ø+pBadMEV, and (C) pLac32EBI+pBadØ metabolomics data were compared. The initial inoculation time point overlaps in PCA space for all samples, so was excluded from plots in the main text for clarity. "O/N" indicates overnight induction; "IND" indicates inoculation induction. All colored ellipses represent 95% confidence intervals for each group.**

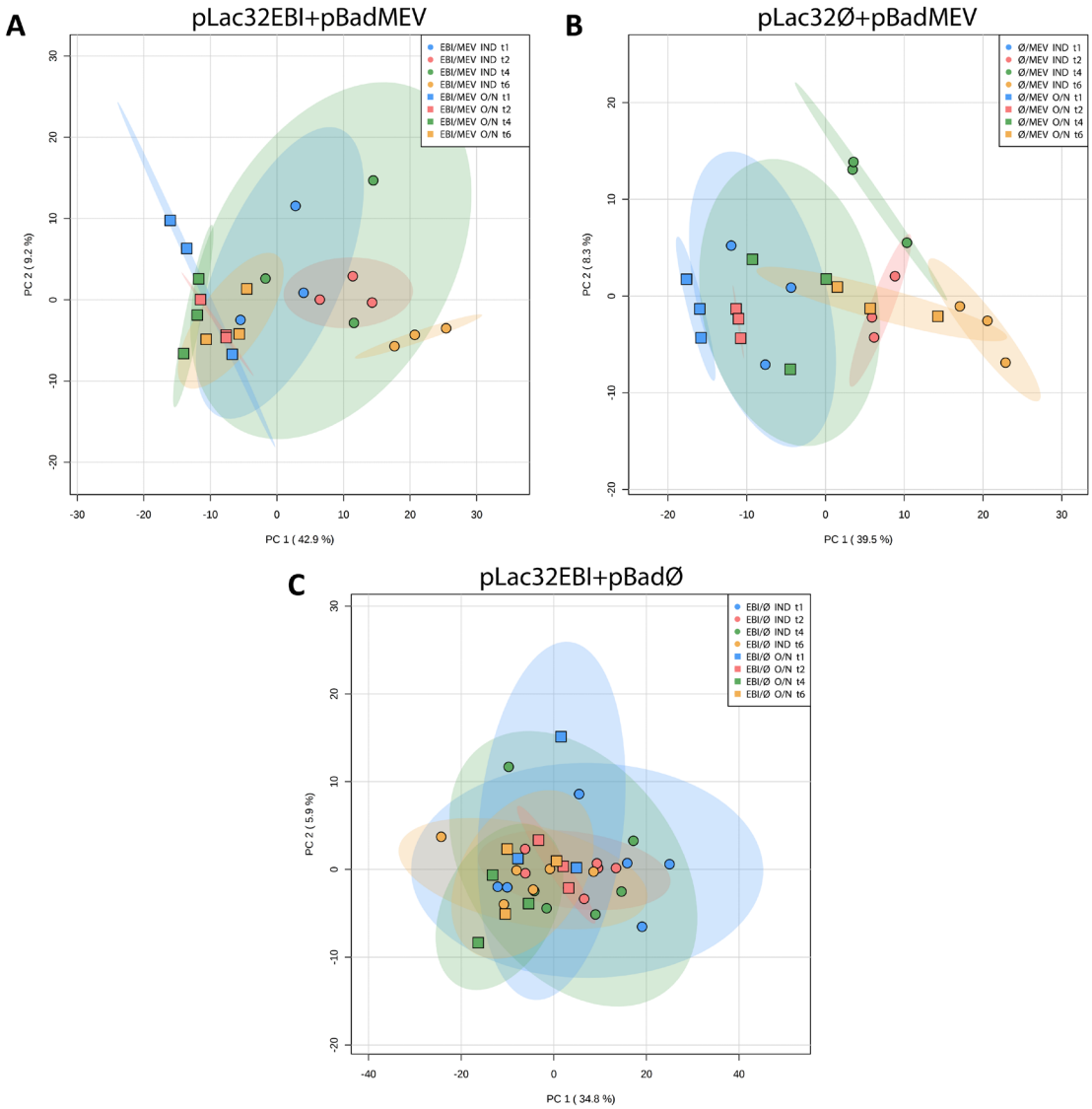


**Figure 12: PCA of overnight- and inoculation-induced strains excluding 0 hours. pLac32EBI + pBadMEV, pLac32Ø + pBadMEV, and pLac32EBI + pBadØ metabolomics data were compared.**
**(A) In pLac32EBI + pBadMEV strains, there is visible separation between induction conditions across the time course, with induction effects far outweighing temporal effects. (B) Separation is also seen for pLac32Ø + pBadMEV strains between induction conditions, though the magnitude of separation is similar to the magnitude of the temporal effects and there is less distinct separation at hour 6. (C) In *pLac32EBI + pBadØ* strains, there is little separation at each time point, indicating no metabolic distinction between induction conditions. "O/N" indicates overnight induction; "IND" indicates inoculation induction. All colored ellipses represent 95% confidence intervals for each group.**

Two-way ANOVA analysis results were generally consistent with what was observed in PCA. In the *pLac32EBI+pBadMEV* strain that completely separated in PCA space based on mevalonate pathway induction, 159 metabolites were significantly different across the time course based on induction time (Table 5). Fewer metabolites were significant for the *pLac32Ø+pBadMEV* strain (only 7, consistent with the decreased separation of induction conditions overall and separation only evident for individual time points). Since there appeared to be a major physiological change in the *pLac32Ø+pBadMEV* strain starting at 4 hours, likely reverting to a state similar to inoculation induction of the mevalonate pathway, we also performed this ANOVA analysis specifically only at hours 1 and 2, as this change may have obscured group effects across the whole time course. In stark contrast to the previous analysis on hours 1 to 6, this analysis identified 60 metabolites with significant induction condition effects. In the *pLac32EBI+pBadØ* strain, only one metabolite was significantly different, consistent with the lack of separation between these two conditions.

**Table 5: Two-way ANOVA for overnight- and inoculation-induced *pLac32EBI+pBadMEV, pLac32Ø+pBadMEV,* and *pLac32EBI+pBadØ*. Contributions of the individual factors are calculated using false-discovery rate (FDR) corrected *p*-values (<0.05).**

| Strain (O/N & IND) | Condition Effects | Time Effects | Interaction Effects | Condition & Time Effects | Time & Interac-tion Effects | Condition, Time, & Interaction Effects |
|---|---|---|---|---|---|---|
| *EBI/MEV* (1-6 hrs) | 159 | 6 | 1 | 9 | 1 | 2 |
| *Ø/MEV (1-6 hrs)* | 7 | 81 | 0 | 71 | 0 | 0 |
| *EBI/Ø (1-6 hrs)* | 1 | 10 | 0 | 0 | 0 | 0 |
| *Ø/MEV (1-2 hrs)* | 60 | 0 | 0 | 3 | 0 | 1 |

### *2.3.6 Time of mevalonate pathway induction significantly affects individual metabolites*

We then analyzed through one-way ANOVA at each individual time point which metabolites were significantly changed across the strains. Similar to the comparison between high and low lycopene producers, no metabolites were found that were obviously related to lycopene synthesis, likely due to limitations on which metabolites are derivatizable and thus detectable via GC-MS. In addition, no metabolites involved in the mevalonate pathway were identified; these analytes should be detectable via GC-MS, so may either be below the limits of detection or not annotated due to the incompleteness of spectral libraries.

Because overnight induction of both *pLac32EBI+pBadMEV* and *pLac32Ø+pBadMEV* caused inhibited growth, for both cases we investigated whether there were changes in the levels of metabolites that have been previously reported in the literature to respond to stress conditions. As discussed above, one well-characterized *E. coli* stress response is the accumulation of amino acids (except for L-alanine and L-methionine, which typically decrease). Almost all amino acid levels in all strains and conditions were identical to those in wild type, with the exception of L-phenylalanine. In inoculation-induced *pLac32EBI+pBadMEV* strains, L-phenylalanine increased over 6 hours and was significantly higher than in overnight-induced *pLac32EBI+pBadMEV* strains and wild type by the end of the experiment (Figure 13A). In inoculation-induced *pLac32Ø+pBadMEV*, L-phenylalanine levels were significantly higher than in overnight-induced and in wild type at multiple time points, though at 6 hours L-phenylalanine levels were actually significantly higher in overnight-induced *pLac32Ø+pBadMEV* than in late induced and wild type (Figure 13B). While the generally higher levels of phenylalanine in

stressed conditions may be relevant, the lack of complete consistency and the lack of similar behavior in other amino acids limits the interpretability of this observation. Another metabolite that has been previously shown to be involved in *E. coli* stress response is N-acetyl-L-alanine. When under heat stress, *E. coli* increases production of this metabolite[116]. Interestingly, N-acetyl-L-alanine had a decreasing trend in both overnight- and inoculation-induced *pLac32EBI+pBadMEV* as well as in wild type, and by hour 2, concentrations in inoculation-induced *pLac32EBI+pBadMEV* were significantly lower than in overnight-induced and wild type (Figure 13C). There were no significant differences in this metabolite in the *pLac32Ø+pBadMEV* and *pLac32EBI+pBadØ* strains. Again, this result suggests that traditional indicators of *E. coli* cellular stress are not evident in the conditions we studied.

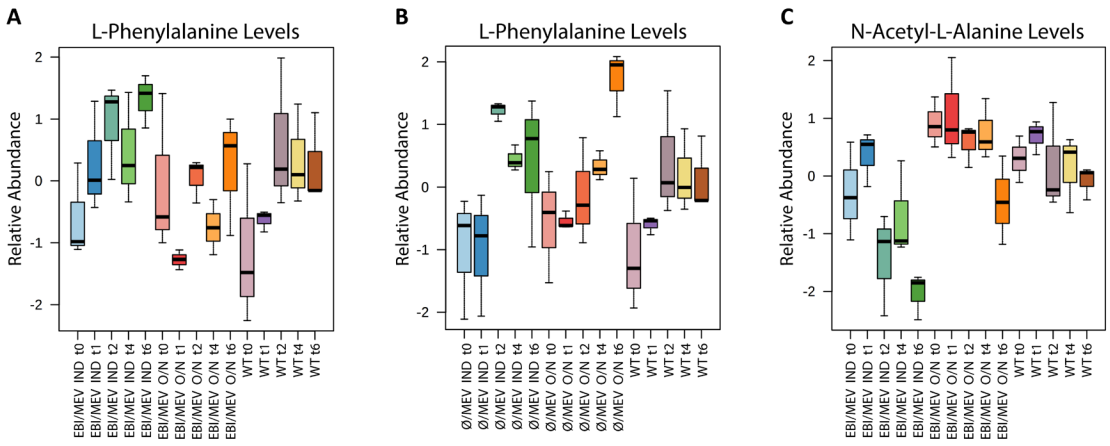


**Figure 13: Profiles of stress-associated metabolites.**
**(A) L-phenylalanine increased over 6 hours and was significantly higher in inoculation-induced *pLac32EBI+pBadMEV* strains than in overnight-induced *pLac32EBI+pBadMEV* strains and wild type by the end of the experiment. (B) L-phenylalanine levels were significantly higher in inoculation-induced *pLac32Ø+pBadMEV* than in overnight-induced or in wild type at multiple time points. However, at 6 hours L-phenylalanine levels were significantly higher in overnight-induced *pLac32Ø+pBadMEV* than in late-induced and wild type. (C) N-acetyl-L-alanine had a decreasing trend in both overnight- and inoculation-induced *pLac32EBI+pBadMEV* as well as in wild type, but to substantially different degrees.**

**By hour 2, concentrations in inoculation-induced *pLac32EBI+pBadMEV* were significantly lower than in overnight-induced and wild type. "O/N" indicates overnight induction; "IND" indicates inoculation induction. Box and whisker plots depict the normalized peak areas. Black lines are the medians, and boxes are the middle 50% values. Error bars represent standard deviation.**

Focusing our analyses on metabolites that were most strongly affected by the time of mevalonate pathway induction, we noticed that homocysteine exhibited interesting trends. For the inoculation-induced *pLac32EBI+pBadMEV* and *pLac32Ø+pBadMEV* strains that showed no growth inhibition, there was an obvious and significant decrease in homocysteine levels over time, with levels much lower than those of wild type (Figure 14A-C). In contrast, the homocysteine levels in the same strains under overnight-induction, growth-inhibited conditions had similar concentrations to those in wild type at each time point (though at 6 hours in the *pLac32Ø+pBadMEV* strain, when cell growth had begun to recover, homocysteine did have a downward trend compared to the earlier time points). Homocysteine in the *pLac32EBI+pBadØ* strain showed a slight downward trend regardless of induction status, consistent with the potential interplay of lycopene and mevalonate pathways' impacts on homocysteine levels.

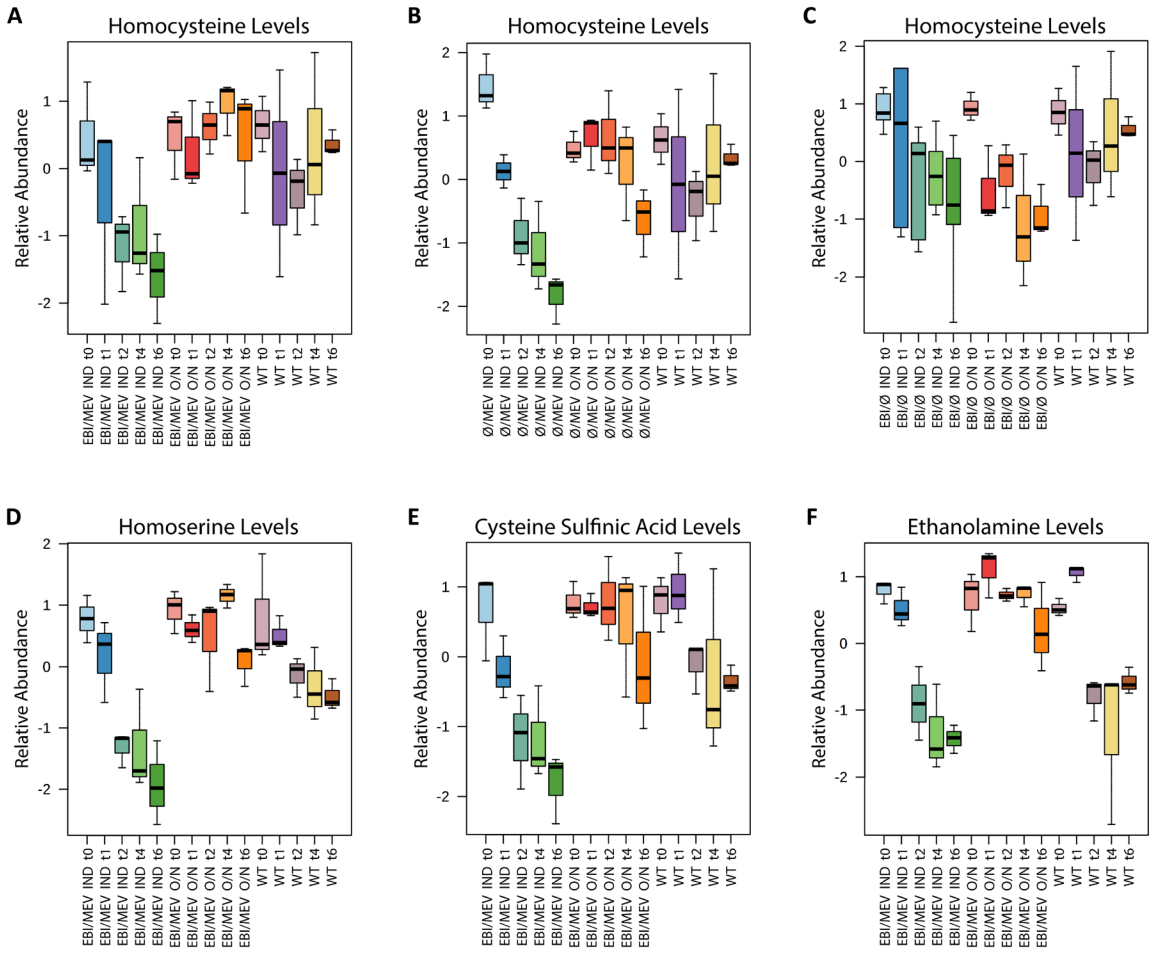


**Figure 14: Levels of metabolites of interest plotted over time.**
**(A) Homocysteine levels for overnight- and inoculation-induced *pLac32EBI* + *pBadMEV* strains have different trends based on induction condition. The overnight-induced strain appears to maintain higher homocysteine levels over the time course than wild type does. (B) Trends similar to this occur in homocysteine levels for *pLac32Ø* + *pBadMEV* strains at each induction condition, except in the overnight-induced strain at hour 6. (C) Homocysteine levels are almost identical at all time points for both overnight- and inoculation-induced *pLac32EBI* + *pBadØ* strains. (D) Homoserine levels in *pLac32EBI* + *pBadMEV* strains have similar trends to the ones seen in homocysteine. (E) Cysteine sulfinic acid and (F) ethanolamine trends are nearly identical to those in homoserine for *pLac32EBI* + *pBadMEV* strains. "O/N" indicates overnight induction; "IND" indicates inoculation induction. Box and whisker plots depict the normalized peak areas. Black lines are the medians, and boxes are the middle 50% values. Error bars represent standard deviation.**

Homoserine, a precursor in homocysteine biosynthesis, also exhibited interesting behavior (Figure 14D). Similar to homocysteine, inoculation-induced pLac32EBI +pBadMEV had lower homoserine levels compared to overnight induction and wild type starting at two hours after inoculation. The downward trend in homoserine was also much stronger in the inoculation-induced strain compared to wild type, with the overnight-induced strain showing no consistent trend with time. In contrast, homoserine levels had no significant changes in *pLac32EBI+pBadØ.*

Two other metabolites had profiles strikingly similar to homoserine in *pLac32EBI+pBadMEV* and wild type (Figure 14E-F): ethanolamine and cysteine sulfinic acid. These metabolites significantly decreased in the inoculation-induced mevalonate- and lycopene-producing strain from hours 2 to 6 but were comparatively constant over the entire 6 hours for the same strain induced overnight. Wild type levels, on the other hand, decreased slightly by hour 2 and remained relatively constant through the rest of the time course. These metabolites did not significantly change in any other strains at either induction condition.

## 2.4 Discussion

In this study, we have used GCxGC-MS metabolomics to characterize the metabolism of engineered lycopene-producing strains of *E. coli*. Previous work by our group showed that changing the RBS on lycopene biosynthesis genes had a significant effect on growth rate and lycopene production. More unexpected, though, was the fact that induction of the mevalonate pathway could have drastically different impacts on lycopene production and cell growth depending on when the pathway was induced[89]. Yet, we had

little knowledge on what metabolic differences, if any, were underlying these changes in behavior.

We surprisingly found that the metabolic impacts of lycopene production in *E. coli* cells are of much smaller magnitude than the metabolic changes inherent to simple batch growth. While we expected the lycopene-producing strains to separate clearly from the wild type strain in PCA plots, we instead found that the first principal component was dominated by temporal variation in metabolite profiles independent of lycopene production. Nonetheless, there were significant differences between the lycopene-producing strains, and analysis of individual time points, in particular the metabolite profiles at 6 hours, showed that the three strains were metabolically distinct. While there were minor fluctuations in L-allothreonine, L-glutamate, and L-phenylalanine levels that have been previously associated with stress response, the behaviors were not consistent enough between strains and across potential stress-indicative metabolites to strongly suggest that the cells are under significant stress when producing lycopene at the levels studied here.

In our previous efforts, we heterologously expressed the mevalonate pathway in the same strain expressing the lycopene pathway, expecting it improve lycopene production because it provides an alternative, non-native path to producing lycopene biosynthetic intermediates, including farnesyl pyrophosphate (FPP). While this was the case when the mevalonate pathway was induced at inoculation at the same time as induction of the lycopene pathway, overnight induction of the mevalonate pathway before inoculation caused lower cell density and smaller increases in lycopene levels, which was quite surprising. These growth profiles suggested that there may have been a toxic intermediate

in the mevalonate pathway that accumulates to growth-inhibiting levels during overnight growth[117]. Our metabolomics results indicate that induction time of the mevalonate pathway has a prominent impact on metabolism, especially when paired with lycopene biosynthesis.

Overnight mevalonate induction clearly has a growth and metabolic impact on cells regardless of whether the cells also express lycopene. When cells do express lycopene, the difference between overnight and inoculation induction of the mevalonate pathway is the most significant source of variation in the data, overwhelming the temporal variability associated with batch growth time course measurements (Figure 12A). This is in itself noteworthy, as changes in metabolism associated just with time-course batch growth are actually greater in magnitude than the differences between strains that have significant vs. zero lycopene production (Figure 5B). It is quite surprising that the impact of whether the mevalonate pathway had been induced in a saturated overnight inoculation culture would be greater than the impact of whether or not cells express a heterologous pathway known to exert significant stress on cell growth and resources.

When cells do not express lycopene, the differences between inoculation and overnight induction of the mevalonate pathway are still striking and easily detectable, though on the same order of magnitude in principal component space as the temporal variations. Moreover, by the end of our sampling period (6 hours), the metabolic profiles of the overnight- and inoculation-induced cultures begin to converge again (Figure 12B), suggesting that the cells are metabolically recovering from the lingering toxicity associated with overnight mevalonate induction. Importantly, this is consistent with the growth

kinetics we observed, where cell growth increased rapidly in the overnight-induced culture starting at 4 hours, with nearly-recovered cell density at 6 hours.

Thus, there is something about lycopene biosynthesis (which, at the levels studied in this paper were minimally toxic and caused only small measurable differences in metabolic profiles) that particularly exacerbates the lingering toxicity associated with overnight mevalonate induction. Without lycopene production, the cells can eventually recover both metabolically and in terms of growth kinetics, but lycopene production is sufficiently antagonistic with the mevalonate pathway to push the cells to an exceedingly large metabolic deviation and prevent any growth recovery.

This analysis also identified two notable metabolites, homocysteine and homoserine, that could play a role in the growth inhibition seen in overnight induction of the mevalonate pathway. These metabolites are precursors to methionine biosynthesis, and are depicted in a metabolic pathway overview diagram along with other metabolites discussed in this paper in Figure 15. Homocysteine has previously been found to inhibit *E. coli* growth when in abundance intracellularly and extracellularly [118, 119]. The drastic decrease seen in homocysteine in inoculation-induced *pLac32EBI+pBadMEV* and *pLac32Ø+pBadMEV* could possibly suggest that these strains, when induced during inoculation, are more equipped to utilize or deplete this metabolite, although the mechanism for such a difference remains unclear. Additionally, the stable trend in homocysteine levels from hours 0 to 4 in the two overnight-induced strains could indicate that early induction of the mevalonate pathway may cause the strains to have a decreased ability to handle homocysteine accumulation, which affects their growth rate. Moreover, the fact that homocysteine levels drop precipitously in the *pLac32Ø+pBadMEV* strain at 6

hours, just as the cells have phenotypically recovered to high growth rates, supports the potential importance of this metabolite in mevalonate-induced toxicity. We note, however, that the wild type homocysteine levels are only slightly lower than the overnight-induced strains, complicating the direct interpretation of the importance of homocysteine levels.

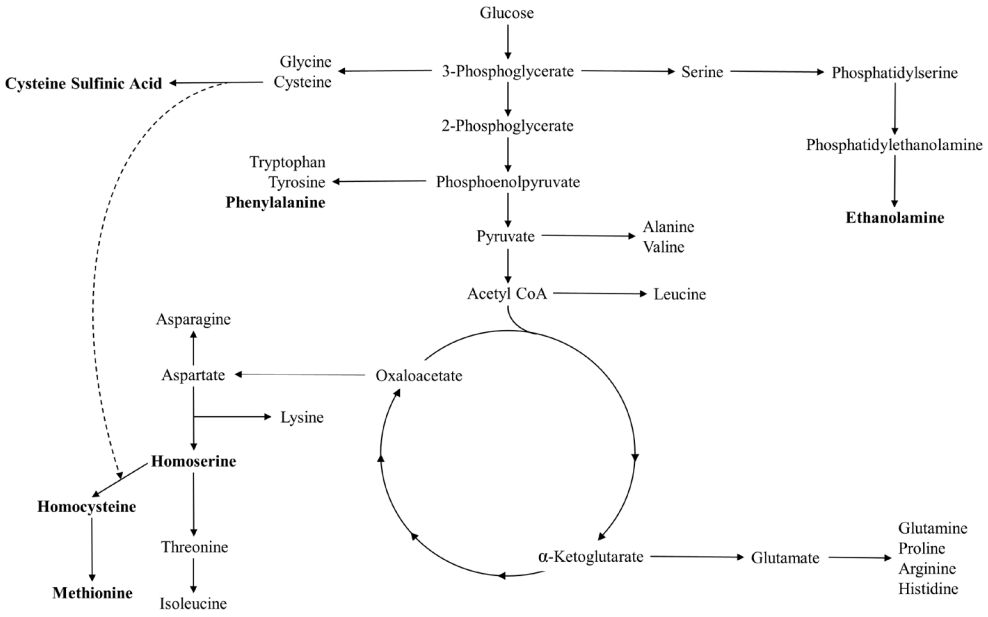


**Figure 15: An overview of metabolic pathways relevant to this work. Metabolites of interest specifically noted in this paper are bolded.**

The same study that reported intracellular homocysteine toxicity also noted that in their attempts to relieve this toxicity, the addition of extracellular homoserine caused intensified inhibition[118]. Our results could then alternatively suggest that the inoculation-induced *pLac32EBI+pBadMEV* strains can better handle homoserine accumulation compared to the overnight-induced counterpart strain. Homoserine's predominantly constant and high concentrations that are only visible in the overnight-induced *pLac32EBI+pBadMEV* strain may contribute to the strain's intensified inhibition even compared to the growth profile of early induced *pLac32Ø+pBadMEV*, which does not have significant variations in homoserine compared to its counterpart strain and wild type.

To test whether homocysteine was contributing to the cytotoxicity seen in overnight-induced *pLac32EBI+pBadMEV* and *pLac32Ø+pBadMEV* strains, we conducted a medium supplementation experiment. Previous work has shown that methionine supplementation can reduce homocysteine-associated toxicity[118]. We thus supplemented the medium with 2mM methionine at inoculation. With addition of methionine, overnight-induced *pLac32EBI+pBadMEV* displayed growth improvements within 2 hours compared to the same strain under the same induction conditions without methionine (Figure 16A). Interestingly, methionine supplementation did not appear to provide significant growth benefits to the *pLac32Ø+pBadMEV* strain (Figure 16B). Although this indicates that homocysteine accumulation is not the sole contributor to the observed toxicity, these results still imply homocysteine contributes to the growth inhibition and likely to the antagonistic effect between the mevalonate and lycopene pathways and is thus likely useful to pursue as a target for further strain engineering.

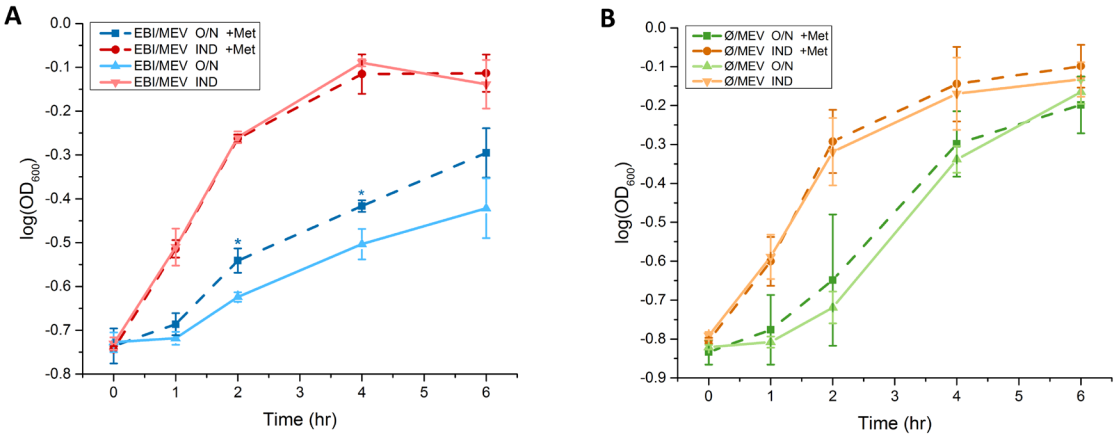


**Figure 16: Growth profiles in response to 2mM methionine supplementation.**
**(A) When the medium is supplemented with 2 mM methionine, the growth inhibition of the *pLac32EBI+pBadMEV* strain induced with arabinose overnight (blue dashed line) is partially alleviated compared to the same strain under the same induction conditions without methionine (blue solid line). Methionine addition has no significant effect on the growth of the *pLac32EBI+pBadMEV* strain induced at**

**inoculation (red dashed and solid lines). (B) Methionine supplementation for overnight-induced (green dashed line) and inoculation-induced (orange dashed line) *pLac32Ø+pBadMEV* strains did not significantly improve growth compared to the same strain under the same induction conditions without methionine (green and orange solid lines), though there does seem to be an insignificant trend towards alleviation at earlier time points which is counteracted by the cells' normal recovery at later time points. "O/N" indicates overnight induction; "IND" indicates inoculation induction; "+Met" indicates supplementation of 2mM methionine. Error bars represent standard deviation. Asterisks indicate statistically significant differences with a *p*-value < 0.05.**

Taken together, these observations suggest the role of a diverse set of metabolites and pathways in the different growth inhibition and metabolic phenotypes we observed. We note, however, that our analysis did not identify as significant many metabolites known to have a direct role in lycopene or mevalonate synthesis. This is likely due in part to limitations in our choice of analytical instrumentation, the GC-MS, as not many such metabolites were even annotated in our dataset. The metabolites in lycopene biosynthesis pathways have few, if any, good leaving groups for derivatization by MSTFA, which would leave them not particularly volatilizable even after derivatization and thus not easily detected by our instrument. Mevalonate pathway metabolites (such as HMG-CoA, IPP, and mevalonate) would be expected to be derivatizable, so if these metabolites are present above the detection limits of our instrument, they may appear as unknown analytes: the metabolite databases used for spectral matching are not necessarily complete for these specific pathways, and we set our identification matching thresholds conservatively to prevent incorrect metabolite annotation.

We also note some inherent limitations in our data. As noted in the Methods section, there were detectable batch effects in the mevalonate induction experiments. We used batch correction software to remove most of the systematic effects in our data, but the batch

effects induced an increase in variability for biological replicates which contributed to a decreased number of analytes with significant time effects in ANOVA analysis and that dampened the visibility of the temporal variation in metabolite profiles. These batch effects are also the likely cause for the decrease in the number of properly aligned and tracked analytes in the induction experiments compared to the initial experiments looking only at different lycopene producers. Nonetheless, trends in individual analytes are consistent across mass spectral acquisition days and across experimental replicate days, supporting the validity of our results.

In addition, our data do not capture quantitative concentrations of metabolites nor identify the mechanisms driving these metabolic divergences. Despite these limitations, our ability to identify fold-changes among metabolites and pathways led us to identify key trends occurring due to heterologous pathway induction. Isotope labeling-based absolute quantification of metabolite levels and systems-scale measurement of gene expression are promising next steps to further elucidate the underlying mechanisms of the trends we identified here. Isotope labeling-based quantification could also help to identify the amount of metabolite leakage during sample quenching to validate that our modified quench protocol provides leakage comparable to the original protocol. Metabolite leakage has not explicitly been tested for and validated in this work; however, since our modified protocol yields a quenched sample with the same ethanol and salt concentrations as the original protocol (just more quickly brought to a cold temperature), we expect the metabolite leakage to be similar. Nonetheless, validation of this hypothesis would help reinforce the broader utility of the modified protocol.

An important caveat in interpreting our results is in our objective, and thus our approach, for strain design and culture. Our primary goal was to engineer strains that produce enough lycopene to visibly turn the cells red in the shortest amount of time possible for a diagnostic readout, not to produce the greatest amount of lycopene in an indefinite timeframe (which often entails significant culture time in stationary phase for non-growth-associated production of lycopene). This is the reason our maximum lycopene production rate of 291.4 ((ng/mL)/OD)/hr from the inoculation-induced *pLac32EBI+pBadMEV* strain is orders of magnitude less than values seen in the literature, which approach 0.030 mg/mL/hr[120-122] over 24 hours of culture. We also note that our experiments were done in the nutrient-rich LB medium. Rich media are commonly used for culturing lycopene producing *E. coli* in order to achieve optimal production rates [90, 121, 122], but rich media may affect nutrient uptake and metabolism, which we had sought to study. However, rich medium is a reasonable model for the final assay mixture in our application, which will contain 25-100% human serum – itself a complex and rich mixture of metabolites. As a result, while these design choices may slightly hinder the generalizability of our results and mechanistic interpretations, they are the most relevant for the target application and the system we will ultimately look to optimize.

In conclusion, we have presented the first profiling of the metabolic differences caused by induction time variation of the mevalonate pathway, explored its potential relationship to lycopene production, identified a possible connection to homocysteine- and homoserine-caused growth inhibition, and validated the involvement of homocysteine-induced toxicity in our system. We also improved the existing metabolomics sampling protocols for *E. coli* cultures to minimize the amount of time cells spent at above-freezing

temperatures that could lead to changes in metabolite profiles. While the underlying mechanisms of the negative effects caused by overnight induction of the mevalonate pathway are still not immediately evident, our efforts have moved us toward a better understanding of the metabolic impacts of this phenomenon and generated hypotheses that could drive future studies. In particular, our work demonstrates the power of metabolomics in helping to provide the understanding needed to drive pathway and strain engineering, and the potential utility of unraveling the mechanisms of homocysteine- and homoserine-related toxicity to allow improved carotenoid biosynthesis.

# CHAPTER 3. INVESTIGATION OF THE METABOLIC CHANGES ASSOCIATED WITH LYSATE PREPARATION METHOD IN CELL-FREE EXPRESSION SYSTEMS

Portions of this chapter are reproduced from my publication "Metabolic Profiling of *Escherichia coli*-based Cell-Free Expression Systems for Process Optimization"[43] in *Industrial & Engineering Chemistry Research*.

## 3.1 Introduction

The previous chapter detailed our efforts to understand the metabolic impacts of optimizing a lycopene reporter system via precursor pathway expression in bacterial biosensors. By applying metabolomics analyses to the various strains and induction conditions, we identified key metabolites involved in the toxicity affecting our engineered strains. Although we were able to present a successful strategy to alleviate some of the stress through metabolite supplementation, our work highlights one of the major limitations of whole-cell biosensor design: metabolite-associated toxicity. Whole cells are inherently limited to the production of compounds, yields, and reaction conditions that can be tolerated by the cell, which can sometimes result in a narrow operating region. Cell-free expression (CFE) systems, which use a protein lysate isolated from cells to produce desired compounds, have the potential to overcome this limitation of whole cells because CFE systems lack a membrane, enabling the detection and production of toxic molecules. Due to the many advantages of CFE systems, they have the potential to expand the scope of

biosensing to entirely new applications, and for these reasons, have been widely pursued in our lab as a biosensing platform.

We have specifically used *Escherichia coli*-based CFE systems, which are the most widely-used and well-characterized CFE systems. *E. coli*-based CFE systems have three main components: protein lysate isolated from *E. coli* cells, a reaction mixture, and a DNA template for gene expression. Activity of the system, generally assessed by yield of a target protein, is heavily dependent on the method used to prepare the protein lysate[39, 123-125]. As mentioned in Chapter 1, lysate preparation protocols consist of growing *E. coli* cells in a rich medium, isolating and washing the cells, lysing the cells, and removing cell debris[40, 123, 126], with the specific details for each step chosen from a variety of potential techniques or reagent combinations. The contents of the growth medium, lysis method, and downstream processing steps vary greatly across different protocols and can dramatically affect the activity of the resulting CFE systems[39], and batch effects are prominent even in lysates that have putatively been identically prepared.

Variation in activity across different lysate preparation methods and the eventual loss of transcriptional and translational functionality of CFE reaction systems after extended reaction periods are often speculatively attributed to a number of potential factors, including (1) protein-level differences in polymerase content or activity and (2) metabolite-level differences comprising the accumulation and depletion of small molecules in a reaction or lysate. However, we are not aware of any large-scale analysis of the small molecules present in different lysates or reactions ever being reported, which is a critical gap in our knowledge, and thus the emphasis of this work. A more complete understanding

of the relationships of small molecules to the activity of CFE systems could be of great value to improve CFE biosensor design and CFE systems overall.

Metabolomics is an ideal tool for characterizing the effects of small molecule differences on CFE activity. Metabolomics analysis could be particularly useful for continuous monitoring of CFE reactions, since small molecules serve as both substrates and energy sources for the enzymes that produce the target compound and other endogenous enzymes that produce by-products detrimental to CFE activity[39, 50]. Thus, metabolic analysis of CFE systems could be used to predict or monitor reaction activity, to ensure batch-to-batch consistency of cell lysate, and even to determine when to add additional compounds to increase and extend the reaction output. However, using metabolomics as a process analysis tool first requires characterizing the baseline metabolic activity of CFE systems and determining which metabolites are most critical for high CFE activity.

Here, we use metabolomics to characterize the effect of lysate preparation method on CFE output. We make lysate from *E. coli* cells using four different standard preparation approaches, characterize their activity and differences in baseline metabolic activity, and then study how incubation time and production of a target protein alter the activity and metabolite profile of each lysate. Using this information, we assess whether supplementation of identified individual metabolites can improve reaction output. Finally, we use our knowledge acquired from the previous experiments to design an improved CFE system.

## 3.2 Experimental Methods

### *3.2.1 Plasmids*

The plasmids pJL1s70 and pJL1 were used in this study. The plasmid pJL1 was a generous gift from Prof. Michael Jewett and Prof. Julius Lucks at Northwestern University. The plasmid pJL1s70 was constructed by replacing the original T7 promoter of pJL1 with a strong $\sigma^{70}$ promoter; the original plasmid is hereafter referred to as pJL1t7. Plasmids were transformed into *E. coli* DH10B cells and isolated with Omega midiprep kits according to the manufacturer's instructions.

### *3.2.2 CFE lysate preparation*

Cellular lysate for all experiments, excluding toehold switch experiments, was prepared based on previously described protocols[39, 40, 123]. To account for potential variability in the preparation process, three batches of each lysate were prepared in parallel, each containing media that was separately prepared and inoculated with individually selected colonies. BL21 cells were grown in either 2x YTP or 2x YTPG media. 2x YTP consisted of 16 g $L^{-1}$ tryptone, 10 g $L^{-1}$ yeast extract, 5 g $L^{-1}$ sodium chloride, 7 g $L^{-1}$ potassium phosphate dibasic, 3 g $L^{-1}$ potassium phosphate monobasic and was pH-corrected to 7.2 with Tris. 2x YTPG media consisted of 75% (v/v) 2x YTP and 25% (v/v) of 7.2% (w/v) dextrose. All media was filter-sterilized prior to use. Cells were grown at 37 °C and 220 rpm to an OD of 2.0, which corresponds with the mid-exponential growth phase. Cells were then centrifuged at 2700 rcf and washed three times with S30A buffer (14 mM magnesium acetate, 60 mM potassium acetate, 10 mM Tris-acetate (pH 8.2), and 2 mM dithiothreitol). After the final centrifugation, the wet cell mass was determined, and

cells were resuspended in 1 mL of S30A buffer per 1 g of wet cell mass. The cellular resuspension was divided into 1 mL aliquots. Cells were lysed using a Q125 Sonicator (Qsonica, Newton, CT) at a frequency of 20 kHz, and at 50% of amplitude. Cells were sonicated on ice with three cycles of 10 seconds on, 10 seconds off, delivering approximately 130 J, at which point the cells appeared visibly lysed. An additional 4 mM of dithiothreitol was added to each tube, and the sonicated mixture was then centrifuged at 12,000 rcf and 4 °C for 10 minutes. The supernatant was removed, aliquoted, and stored at -80 °C for future use or subject to a runoff reaction and dialysis. For the runoff reaction, 1 ml aliquots of lysate were incubated in 14 ml round bottom culture tubes at 37 ˚C and 220 rpm for 80 minutes. After this runoff reaction, the cellular lysate was centrifuged at 12,000 rcf and 4 ˚C for 10 minutes. The supernatant was removed and loaded into a 10 kDa MWCO dialysis cassette (Thermo Fisher). Lysate was dialyzed in 1L of S30B buffer (14 mM magnesium glutamate, 60 mM potassium glutamate, 1 mM dithiothreitol, pH-corrected to 8.2 with Tris) at 4 ˚C for 3 hours. Dialyzed lysate was removed and centrifuged at 12,000 rcf and 4 ˚C for 10 minutes. The supernatant was removed, aliquoted, and stored at -80 ˚C for future use. Total protein concentration of each lysate batch was assessed with a Bradford assay.

### *3.2.3 CFE reactions*

Cell-free reactions for all experiments were run as previously described[127]. Each cell-free reaction contained 0.85 mM each of GTP, UTP, and CTP, in addition to 1.2 mM ATP, 34 μg/mL of folinic acid, 170 μg/mL *E. coli* tRNA mixture, 130 mM potassium glutamate, 10 mM ammonium glutamate, 12 mM magnesium glutamate, 2 mM each of the 20 standard amino acids, 0.33 mM nicotine adenine dinucleotide (NAD), 0.27 mM

coenzyme-A (CoA), 1.5 mM spermidine, 1 mM putrescine, 4 mM sodium oxalate, 33 mM phosphoenol pyruvate (PEP), 27% cell lysate, and 12 nM of the specified plasmid. For each experiment, a fresh aliquot of lysate was used to minimize variability caused by multiple freeze-thaw cycles, and all experiments were performed within four months of lysate preparation. When specified, the chemicals putrescine, spermidine, homocysteine, β-alanine, glucose 6-phosphate disodium salt hydrate, and sodium chloride were added to reactions at the specified concentrations. All chemicals were purchased from Sigma Aldrich.

For metabolomics analysis, 210 μL reactions were prepared in 2 mL microcentrifuge tubes. Reactions were run in biological triplicate, with each replicate containing a different batch of lysate. Samples were incubated at 37 °C for the specified time. 10 μL of the reaction was then removed and stored at -80 °C for subsequent fluorescence analysis, and the remaining 200 μL was stored at -80 °C for subsequent metabolomics analysis. In experiments solely assessing GFP production, 12 μL reactions were prepared in PCR tubes and incubated at 37 °C for the specified time.

### *3.2.4 Assessment of GFP production*

To measure fluorescence of CFE reactions, 8 μL of each reaction was added to a well of a 384 well plate, and the fluorescence was measured with a plate reader (Synergy4, BioTek). Excitation and emission for GFP were 485 and 510 nm, respectively.

### *3.2.5 Protein precipitation for metabolomics analysis*

Before beginning the protein precipitation protocol, 20 µL was removed from all samples to prepare pooled quality control (QC) samples for the experiment comparing the lysates before reaction mixture and DNA addition, the experiment comparing reaction mixtures after protein synthesis, and the experiment comparing reactions after expression driven by T7 and $\sigma^{70}$ promoters, respectively. For the in-depth time course analysis of reactions using lysate NN, 16 µL were removed from all samples for a pooled QC sample. These pooled QC samples were directly prepared with all other lysate samples for protein precipitation.

Proteins were precipitated from all samples stored for metabolomics analysis via the following protocol: first, methanol was added to each sample at a 1:2 sample to methanol ratio and vortexed briefly. The samples were incubated at -20 °C for 20 min, centrifuged at 11,600 rcf for 30 min at room temperature, and the supernatant was collected. The supernatants of the pooled QC samples were then evenly aliquoted in to three, six, and two tubes for the experiment comparing reactions after protein synthesis, the experiment comparing reactions after expression driven by T7 and $\sigma^{70}$ promoters, and the in-depth time course analysis of reactions using lysate NN, respectively. For the experiment comparing lysates before reaction mixture and DNA addition, one QC tube was used. The supernatants were dried at 40 °C in a CentriVap until all water was removed and stored at -80 ºC.

### *3.2.6 GC-MS analysis*

Before derivatization, the lysate samples and QC samples were transferred to a CentriVap to be dried at 40 °C for 15 min. Samples were derivatized as previously

described[99, 128]. 10 µL of 20 mg/mL O-methylhydroxylamine hydrochloride (MP Biomedicals, LLC, Santa Ana, CA, United States) in pyridine was added to each dried sample and shaken at 1400 rpm for 90 min at 30 °C. 90 µL of N-methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) + 1% trimethylchlorosilane (TMCS) (Thermo Scientific, Lafayette, CO, United States) was then added to the samples and shaken at 1400 rpm for 30 min at 37°C. Samples were centrifuged at 21,100 g for 3 min, and 50 µL of the supernatant was added to an autosampler vial. Samples were spiked with 0.25 µL of a retention time standard solution composed of fatty acid methyl esters (FAMES). At the beginning of the GC-MS run, the QCs were injected once and repeated again after every 4-6 sample injections to allow for downstream correction for batch effects. A derivatization blank was prepared and run with every batch of samples.

A LECO Pegasus 4D instrument with an Agilent 7683B autosampler, Agilent 7890A gas chromatograph and time-of-flight mass spectrometer (TOF-MS) was used to analyze the samples. The first column was an HP-5, 28 m long × 0.320 mm ID × 0.25 µm film thickness (Agilent, Santa Clara, CA, United States), and the second was an Rtx-200, 1.7 m long × 0.25 mm ID × 0.25 µm film thickness (Restek, Bellefonte, PA, United States). More detailed gas chromatography, autosampler, and mass spectrometry methods are provided in Appendix A.

### *3.2.7 Data analysis*

Sample runs were analyzed in ChromaTOF (LECO, St. Joseph, MI, United States) to determine baseline, peak area, and peak identification as described previously.[101, 102] Briefly, settings included a baseline offset of 0.5, automatic smoothing, 1st dimension peak

width of 36 s, 2nd dimension peak width of 0.10 s, and a match of 700 required to combine peaks with a minimum signal-to-noise (S/N) of 5 for all subpeaks. Peaks were required to have a S/N of 10 and have a minimum similarity score of 800 to NIST, Golm and in-house spectral libraries before assigning a name. Unique mass was used for area and height calculation.

MetPP was used to align the samples.[129] Sample files and a derivatization reagent blank file were uploaded from ChromaTOF. Unknowns were retained during the peak alignment process. The derivatization reagent blank file was used to subtract peaks resulting from the sample preparation reagents from the corresponding sample files. On-the-fly alignment was used with manually selected quality control samples as the peak list for primary alignment. Peak alignment was performed using the default criteria.

For the T7 and $\sigma^{70}$ promoter comparison experiment, one of the samples from the group with no plasmid and no T7 RNA polymerase was not properly injected into the GC-MS. For this group, we took the average of the abundances of the two properly injected samples, so there would be three samples in the group for downstream analyses.

To remove analytes that were not reproducibly detected, analytes for which more than half of the values were missing in the QC samples or for which the QC samples had a coefficient of variance larger than 0.5 were removed from the data set. Then, missing values were manually corrected using small value correction only if all the values were missing in the biological replicates.

After alignment, further processing of the data was done to the samples for the in-depth time course analysis of reactions using lysate NN based on the procedure to remove

evident batch effects. First, batch effects were reduced from the data set using LOESS, as previously described[130]. The data set was then processed with Combat[131] to further remove batch effects that were evident from principal component analysis of the LOESS-corrected data.

Finally, MetaboAnalyst was used for statistical and pathway analysis[132]. For both analyses, remaining missing values were k-nearest neighbors (KNN) corrected. Data was filtered using the interquantile range method and then log-transformed using generalized logarithm transformation (base 2) and autoscaled. P-values were adjusted using the Benjamini-Hochberg False Discovery Rate (FDR). Differences were considered significant at FDR-corrected $p < 0.05$. For pathway analysis, the *Escherichia coli* K-12 MG1655 (KEGG) pathway library was selected using a global test for pathway enrichment analysis and relative-betweenness centrality for pathway topology analysis.

For the experiment evaluating the effects of metabolite supplementation, the impact of the interaction of metabolites on GFP output was estimated using the Regression tool in the Data Analysis package of Excel.

The metabolomics datasets for this study are available via the Metabolights repository, with the dataset identifier MTBLS1079.

## 3.3 Results and Discussion

### *3.3.1 Lysate preparation method affects CFE activity and metabolite profile*

We first aimed to characterize the effect of lysate preparation on both the activity of CFE systems and the baseline metabolite profile of the protein lysates. We prepared four

lysates using previously published preparation techniques (Figure 17A). *E. coli* cells were grown either in a growth medium that contained glucose (lysates GN and GD) (signified by G as the first letter, for glucose) or in a medium without added glucose (lysates NN and ND) (signified by N as the first letter, for no glucose). Cells were isolated during exponential growth and lysed via sonication, and the protein lysate was then removed from the cell debris. The resulting lysate was either subjected to further processing, specifically a "run-off" reaction and a four-hour dialysis (lysates ND and GD) (signified by D as the second letter, for dialysis) into a glutamate buffer or stored for future use (lysates NN and GN) stored for future use (signified by N as the second letter, for no dialysis).

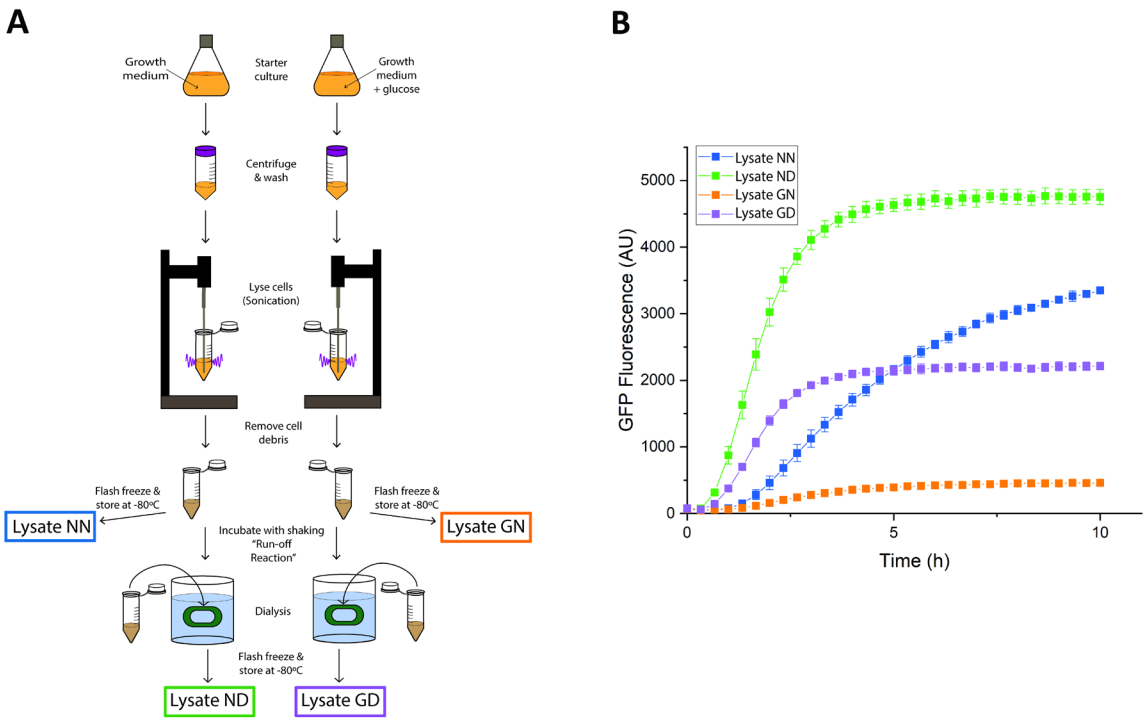


**Figure 17: Lysate preparation overview and GFP output.**
**(A) Lysates GN and GD and lysates NN and ND were collected from cells cultured in growth media with or without glucose, respectively (signified by G or N as the first letter, for glucose or no glucose, respectively). After centrifugation and washing, the cells were lysed via sonication and cellular debris was removed. At this point, lysates ND and GD underwent a "run-off reaction" and dialysis before storage at -80°C**

**(signified by D as the second letter, for dialysis) and lysates NN and GN were stored at -80°C (signified by N as the second letter, for no dialysis). (B) Lysate preparation strongly affects the output of CFE reactions. Reactions using each of the four lysates were assembled and used to produce GFP. GFP production was monitored by measuring fluorescence across 10 hours. Error bars represent standard deviation of biological triplicates.**

Consistent with previously reported data, we found that preparation method strongly affects lysate ability to produce a target protein. Because we aimed to capture the activity of native lysate transcription and translation, we added as a DNA template the plasmid pJL1s70, which uses a standard *E. coli* transcriptional promoter to control expression of green fluorescent protein (GFP). Lysates prepared from cells grown in the presence of glucose produced less GFP than their counterparts grown in the absence of glucose (Figure 17B), and dialysis increased expression in both cases. The best-performing lysate both in terms of protein production rate and final protein expression was the lysate that was isolated from cells grown without glucose and then dialyzed (lysate ND).

Because all of the cells were isolated in mid-exponential phase and the lysates contained the same levels of total protein (Figure 18) yet they still produced different levels of GFP, we hypothesized that there may be differences in metabolite levels between the lysates. We used metabolomics to analyze the metabolic profile of each lysate (before any reaction was performed with the lysate). We precipitated the proteins and then analyzed the remaining metabolite mixture with two-dimensional gas chromatography coupled to mass spectrometry (GCxGC-MS). Processing the resulting instrument output with a standard metabolomics workflow (see Experimental Methods) yielded relative abundances for 260 identified and unannotated analytes that were used for downstream data analyses.

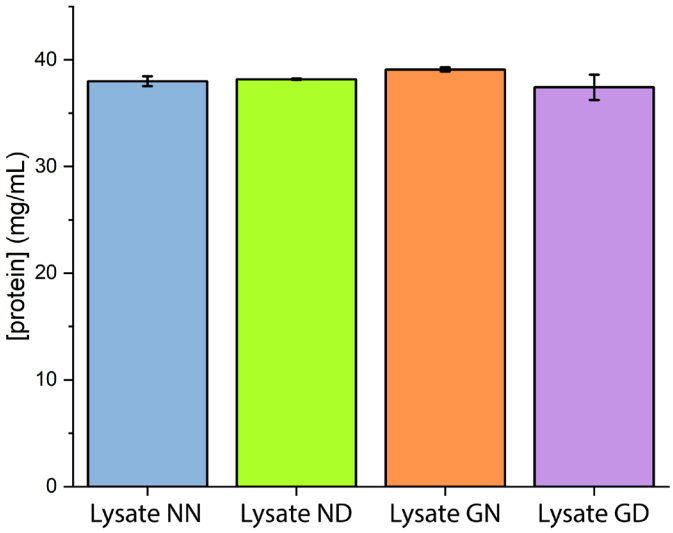


**Figure 18: Total protein concentration in each lysate.**
**Lysates have essentially the same protein concentration. Error bars represent standard deviation of biological triplicates.**

To assess differences in metabolite profiles of the four lysates, we first analyzed the data with principal component analysis (PCA). an unsupervised dimensional reduction approach that identifies the modes of the data (linear combinations of analyte measurements) that capture the most variance, which are referred to as principal components. Separation of sample groups on the resulting PCA plots indicates substantial differences in overall metabolite profiles between the different lysates. Dialysis has the strongest impact on total metabolite profiles, apparent by the complete separation of dialyzed and non-dialyzed lysates in the first principal component (Figure 19A). During dialysis, small molecules diffuse through a semi-permeable membrane into the external buffer, decreasing the overall concentration of small molecules within the lysate. Though dialysis accounts for most of the variation in metabolite profile, assessment of the second and third principal components shows that the media composition is also a source of variation, apparent by the clustering of samples based on growth media (Figure 19B).

Univariate analysis via analysis of variance (ANOVA) supported these conclusions, with dialysis having a significant effect on almost half of the measured analytes; a small number of other metabolites exhibited effects due to glucose in the growth medium or due to interactions between the two factors (Figure 20). The clear differences in the metabolic profiles of different lysates thus clearly demonstrate that preparation protocol strongly affects both lysate composition and potential for protein production, and they provide a potential baseline for future quality control of any individual lysate preparation method.

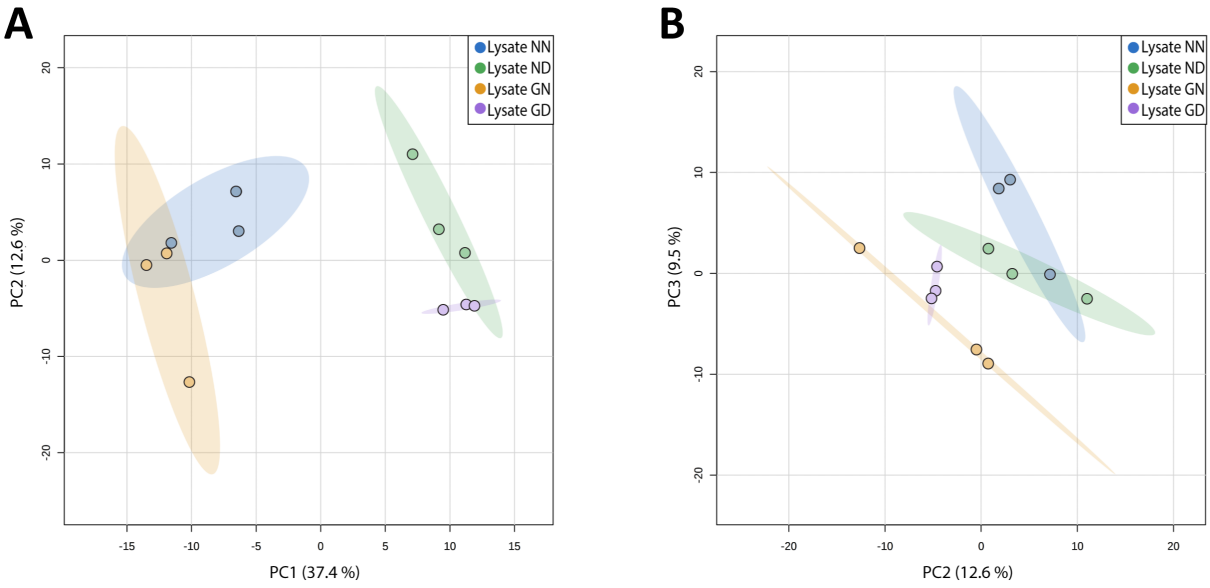


**Figure 19: PCA of metabolomics data collected from all lysates.**
**Lysates (without reaction mixture or plasmid) are visualized in (A) principal components 1 and 2 and (B) principal components 2 and 3. The dialyzed lysates (lysates ND and GD) separate from the non-dialyzed lysates (lysates NN and GN) in principal component 1, whereas lysates derived from starter cultures grown without glucose (lysates NN and ND) separate from lysates prepared from cultures grown with glucose (lysates GN and GD) in principal components 2 and 3. Colored ellipses represented 95% confidence intervals for each group, and the plotted samples are biological replicates.**

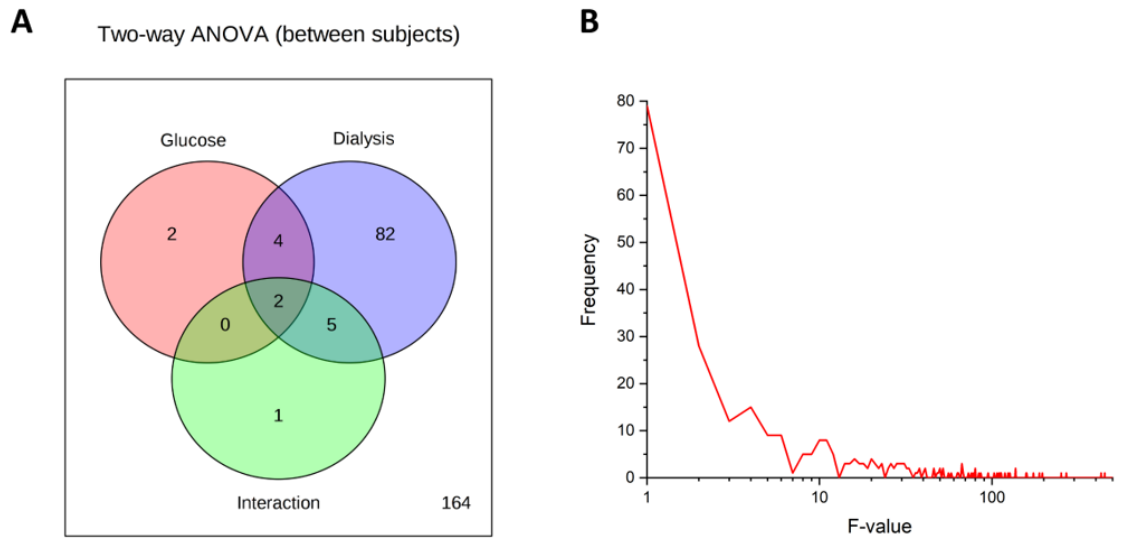


**Figure 20: Analysis of the effects of glucose in the culture medium and the dialysis of lysates on the metabolite profile.**
**(A) Two-way ANOVA and (B) the distribution of f-values of metabolomics data collected from all lysates (without reaction mixture or plasmids), assessing the effects of the presence of glucose in the culture medium and the dialysis of lysates. Dialysis has the strongest effect on the metabolic differences between lysates. In (a), numbers represent counts of metabolites that have significant contributions to the individual factors, which are assessed using false-discovery rate (FDR) corrected *p*-values (<0.05). In (B), bin values of 1 were used to determine f-value frequency. The majority of f-values are between 0 and 1.**

### *3.3.2 Baseline lysate activity accounts for most metabolic changes*

While analysis of the starting lysates may yield benefits for quality control of lysate production, a better understanding of the ways that metabolite profiles change during CFE reactions could ultimately be used to design online reaction monitoring and control strategies, or to rationally determine protocol modifications that could extend reaction time. To that end, we next characterized the temporal changes in metabolite levels associated with protein synthesis. For each lysate, we assembled CFE reactions with and without a DNA template for protein synthesis (the plasmid pJL1s70), with the goal of decoupling the baseline changes in metabolites that occur even without protein production from the changes specifically caused by synthesis of the target protein. We characterized both the

initial metabolic state of the system and its final metabolic state after protein production plateaued at 8 hours. After all data processing, these metabolomics measurements yielded relative abundances for 284 identified and unannotated analytes that were used for further analyses.

Though reactions without plasmid produce no GFP (Figure 21), their metabolic profiles as visualized using multivariate analysis are strikingly similar to the corresponding reactions containing plasmid (Figure 22A-D). Surprisingly, incubation time (rather than protein synthesis) causes the most prominent metabolic changes and is thus the phenomenon captured by the first principal component. Univariate analysis via two-way ANOVA further supports the finding that incubation time, rather than protein expression, accounts for the majority of metabolic changes within a CFE reaction (Table 6), though each lysate contains a number of analytes with statistically significant effects due to protein expression (Figure 23).

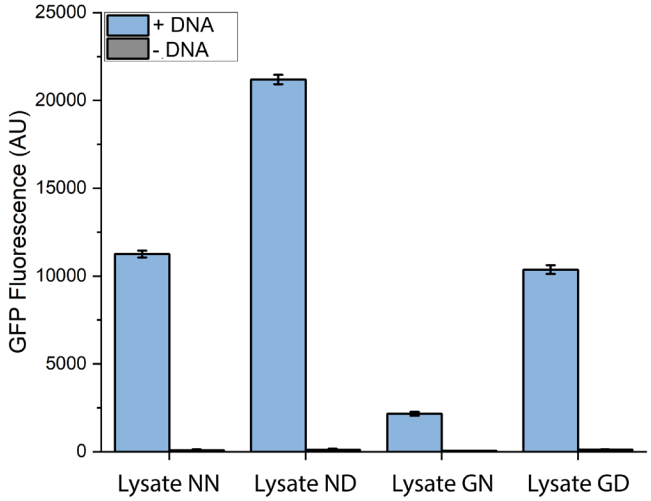


**Figure 21: GFP fluorescence data of all lysates after 8 hours of GFP synthesis. Error bars represent standard deviation of biological triplicates.**

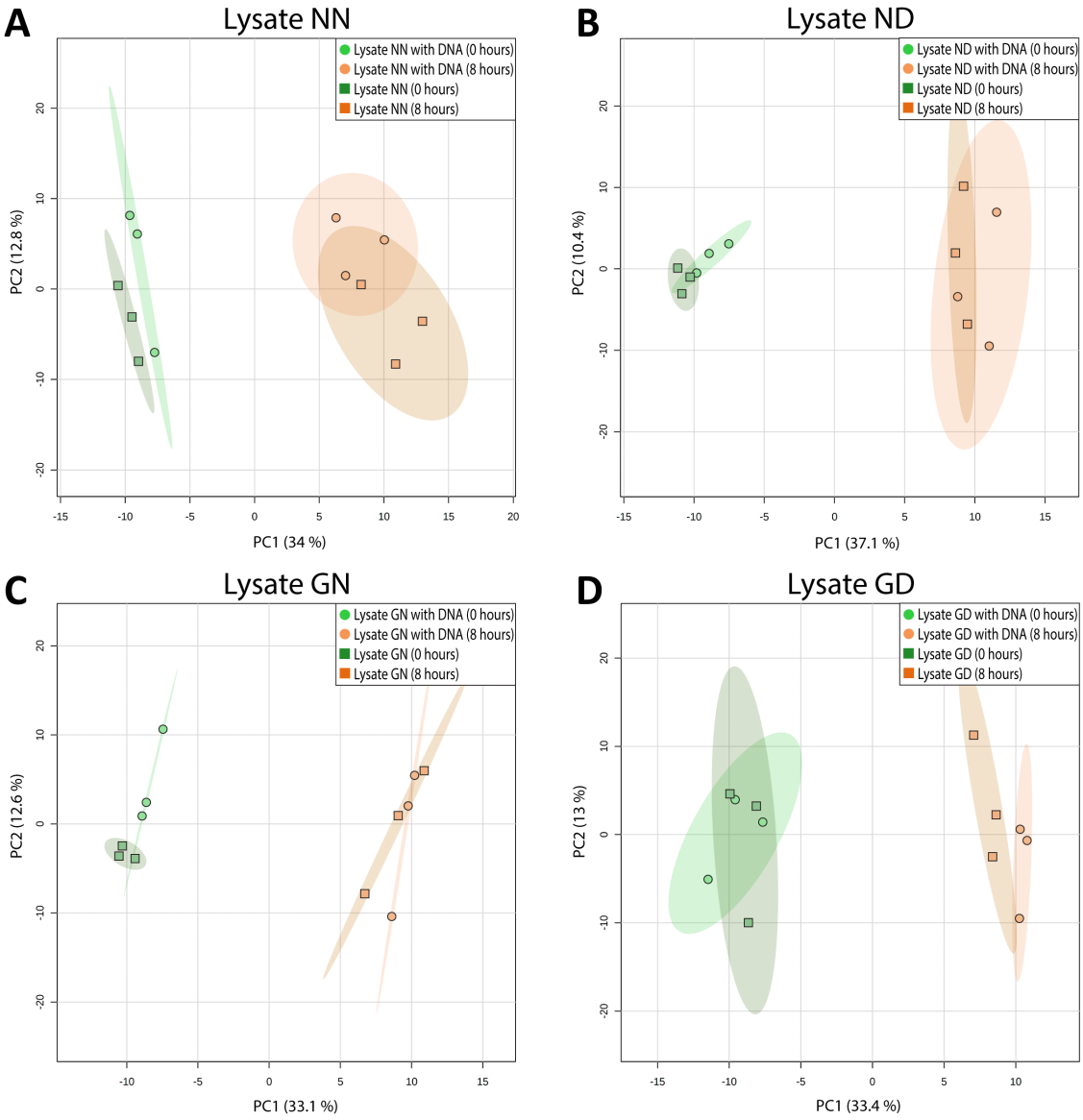


**Figure 22: PCA of metabolomics data collected from CFE reactions using each lysate. Reaction samples using (a) lysate NN, (b) lysate ND, (c) lysate GN, and (d) lysate GD were collected at the initial time point and after 8 hours of incubation. Each lysate exhibits distinct separation based on time, but little separation based on plasmid presence or absence. Colored ellipses represent 95% confidence intervals for each group, and the plotted samples are biological replicates.**

**Table 6: Two-way ANOVA of metabolomics data collected from CFE reactions in all lysates.**

| Lysate | Plasmid Effects | Time Effects | Interac-tion Effects | Plasmid & Time Effects | Plasmid & Interac-tion Effects | Time & Interac-tion Effects | Plasmid, Time, & Interaction Effects |
|---|---|---|---|---|---|---|---|
| Lysate NN | 9 | 95 | 8 | 27 | 16 | 18 | 82 |
| Lysate ND | 1 | 83 | 1 | 4 | 0 | 0 | 12 |
| Lysate GN | 0 | 239 | 0 | 2 | 0 | 0 | 12 |
| Lysate GD | 0 | 66 | 1 | 3 | 0 | 2 | 10 |

*Numbers represent counts of metabolites that have significant contributions to the individual factors, which are assessed using false discovery rate (FDR)-corrected p-values (<0.05).*

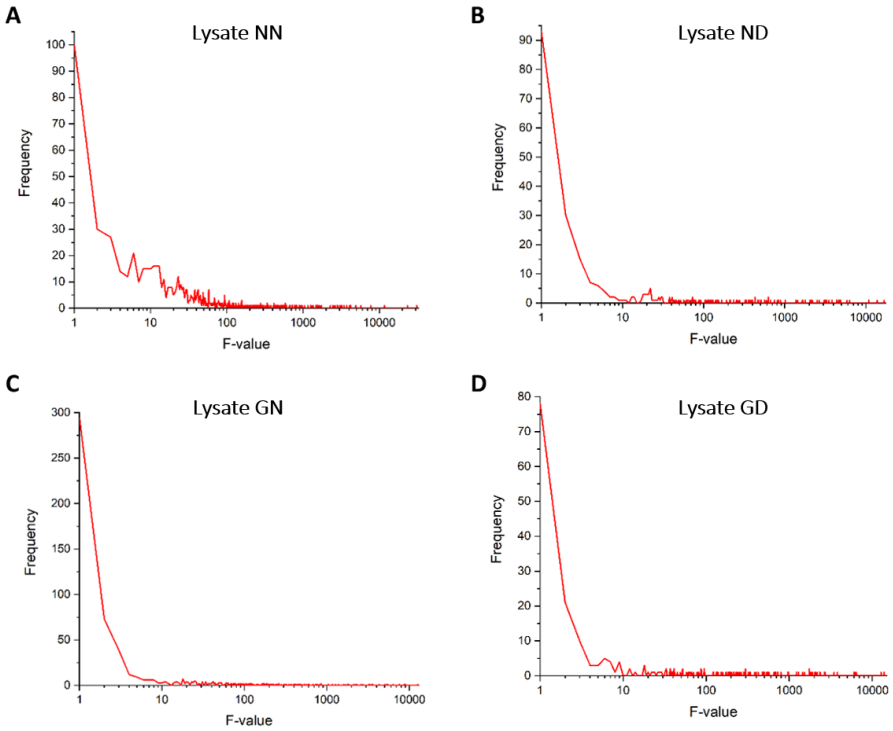


**Figure 23: F-value distribution graphs for two-way ANOVA data for each lysate. (A) Lysate NN, (B) ND, (C) GN, and (D) GD are evaluated after 8 hours of GFP production. The majority of f-values are between 0 and 1 for all lysates. Bin values of 1 were used to determine f-value frequency.**

To assess whether metabolic changes driven by baseline metabolic activity or target protein expression could be occurring earlier in the reaction, we performed a higher resolution time course analysis for one of the lysates. We selected lysate NN for this analysis because it was a fairly strong GFP-producing lysate, we expected it to have more diverse metabolic content since it had not been dialyzed, and it produced its protein gradually over the course of 8 hours such that we might be better able to capture temporal variation in protein expression-dependent changes in metabolites. We set up reactions using this lysate with and without a plasmid DNA template for 0, 1, 2, 4, and 8 hours. After all data processing, these metabolomics measurements yielded relative abundances for 387 identified and unannotated analytes that were used for further analyses. Multivariate analysis of the results via PCA shows that incubation time, rather than GFP production, accounts for the majority of metabolic variation (Figure 24), in agreement with our previous results. Univariate analysis via two-way ANOVA further supported this conclusion (Figure 25). It is thus clear that metabolic activity in cell-free reactions is dominated by contributions inherent to the reaction components alone and with comparatively weak dependence on protein expression, a finding that complicates efforts to design process monitoring and control approaches using small molecule indicators.

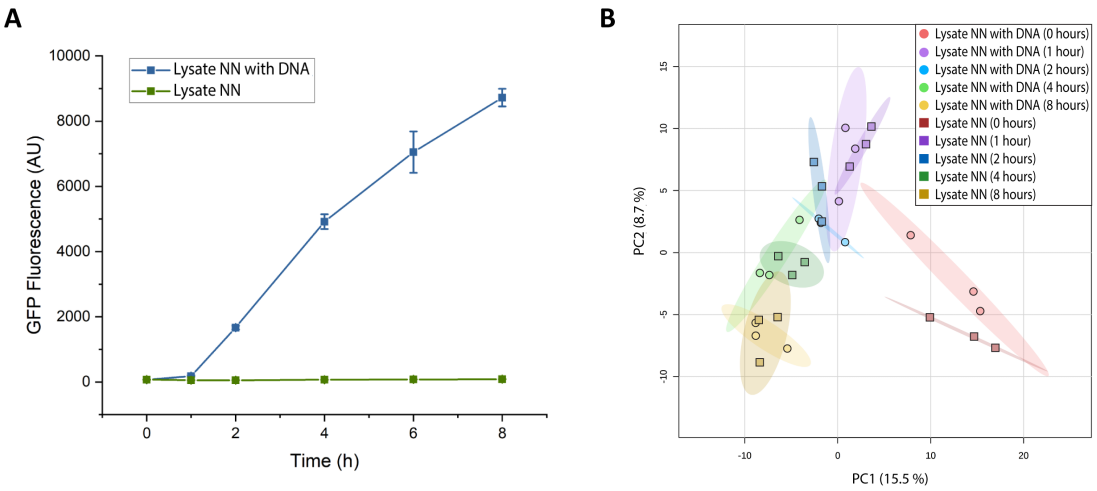


**Figure 24: Protein production and metabolomics analysis of CFE reactions using lysate NN.**
**(A) GFP production and (B) PCA of metabolomics data from CFE reactions run in lysate NN, either with or without the pJL1s70 plasmid. Samples were collected over the course of 8 hours. PCA shows that samples with and without the pJL1s70 plasmid overlap at each timepoint. Colored ellipses represent 95% confidence intervals for each group. Error bars represent standard deviation of biological triplicates.**

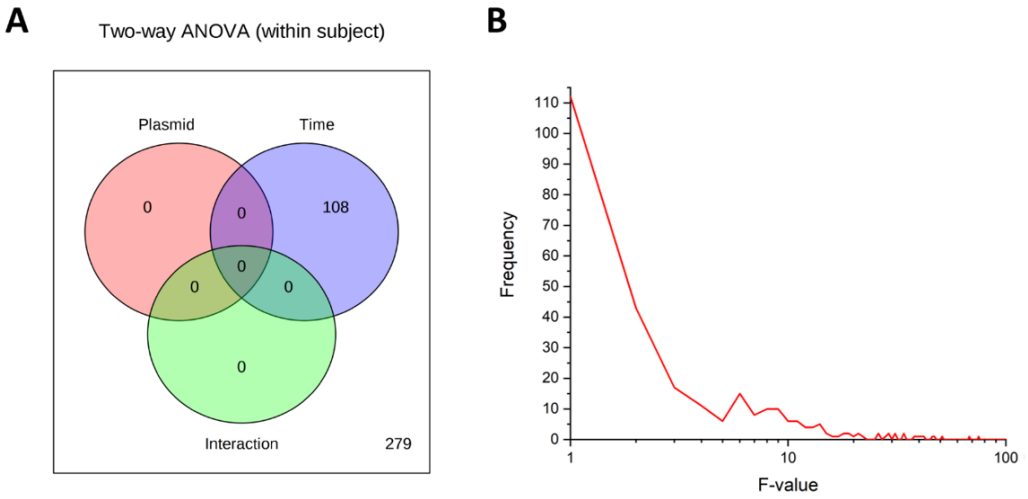


**Figure 25: Two-way ANOVA and the distribution of f-values of metabolomics data collected from CFE reactions using lysate NN.**
**Samples were collected over the course of 8 hours to assess the effects of reaction time and plasmid addition. In (A), numbers represent counts of metabolites that have significant contributions to the individual factors, which are assessed using FDR corrected p-values (<0.05). In (B), bin values of 1 were used to determine f-value frequency. The majority of f-values are between 0 and 10.**

To further test the theory that endogenous metabolic activity overwhelms metabolic changes caused by protein production, we analyzed the changes of metabolite levels in CFE systems with much higher protein production. We ran reactions using the T7 transcriptional system, an orthogonal system that can produce up to an order of magnitude more target protein than systems that rely on endogenous *E. coli* transcription. We added T7 RNA polymerase and an appropriate DNA template (the plasmid pJL1t7, which contains GFP under control of the T7 promoter) to reactions in each of the four lysates and showed that protein expression was dramatically increased in all reactions, with the most prominent effects in reactions run with lysate GD (Figure 26). We again performed metabolomics and analysis, and after all data processing, these metabolomics measurements yielded relative abundances for 418 identified and unannotated analytes that were used for further analyses. Metabolomics analysis of lysate GD with this transcriptional system revealed that after 8 hours of reaction time, there is no discernible pattern of systems-scale metabolic changes for extremely high protein expression compared to reactions that produce no GFP (Figure 27). Multidimensional analysis via PCA shows no apparent separation of samples with the plasmid from samples without plasmid, suggesting that the changes in the metabolic profile of CFE reactions due to even very high levels of protein expression are dominated by changes caused by endogenous metabolic activity.

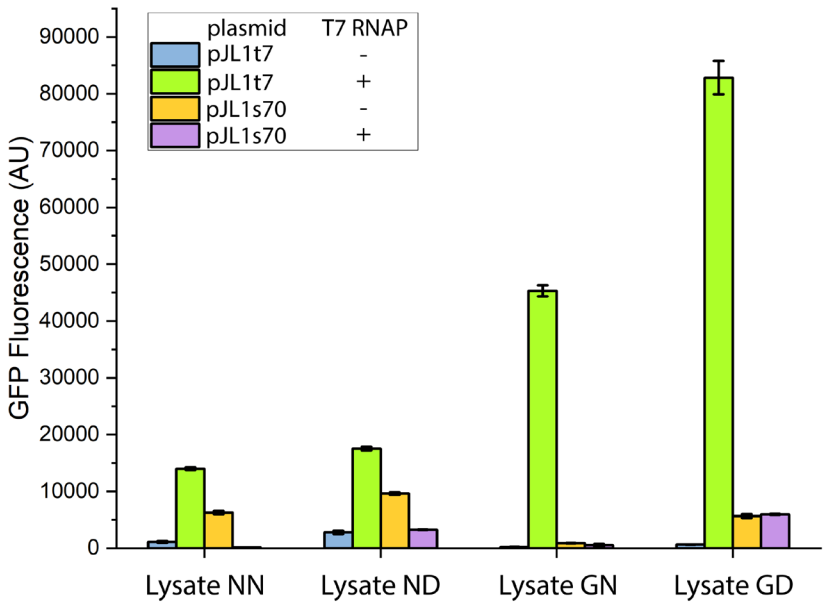


**Figure 26: GFP production from different expression systems in reactions run in different lysates.**
**GFP expression was controlled by either a T7 promoter (pJL1t7) or a standard E. coli $\sigma^{70}$ promoter (pJL1s70), and T7 RNA polymerase was added as indicated. Fluorescence was measured after 8 hours of incubation. Error bars represent standard deviation of biological triplicates.**

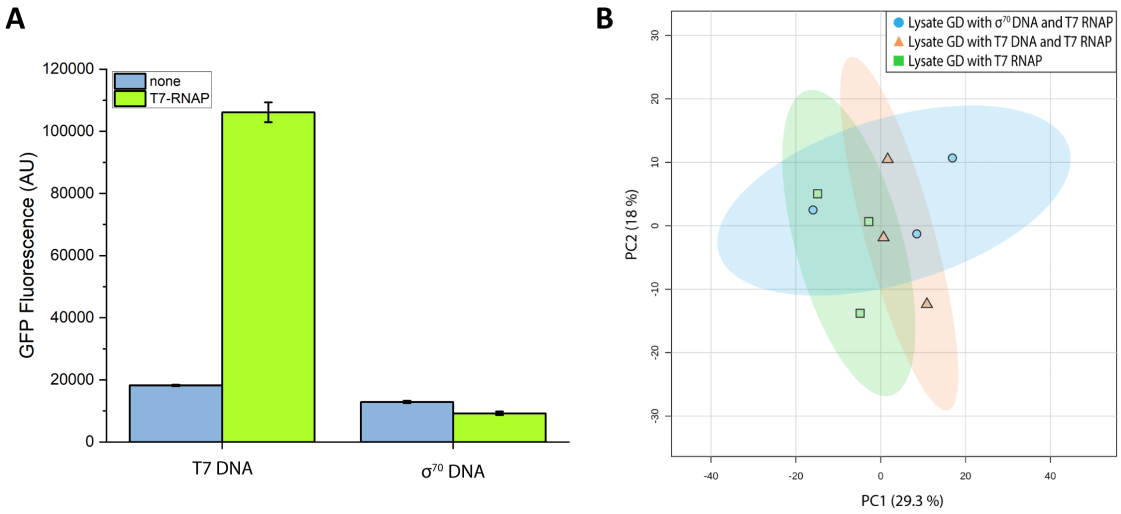


**Figure 27: Characterization of T7- vs. $\sigma^{70}$-driven transcription of GFP in lysate GD.**
**(A) T7-driven transcription of GFP yields over 4 times more GFP than $\sigma^{70}$-based expression. (B) PCA of the metabolomics data corresponding with these reactions shows little separation between reactions that contain pJL1t7, pJL1s70, and no plasmid. Error bars represent standard deviation of biological triplicates. Colored ellipses represented 95% confidence intervals for each group.**

### *3.3.3 Endogenous lysate metabolism affects protein-production capacity*

Because metabolic activity in CFE reactions appears to be highly coupled to endogenous lysate metabolic activity, we hypothesized that the metabolic state of the lysate could dictate the protein production capacity of the reaction. We further hypothesized that consumption of reaction mix reagents by native lysate enzymes or accumulation of potentially toxic compounds could decrease the ability of lysates to produce the target protein. To test this, we assembled CFE reactions with everything except a DNA template, incubated the reactions for eight hours, and then subsequently added plasmid DNA template to determine if the pre-incubated reactions could still produce GFP. When reactions were incubated at 37 ºC prior to plasmid addition, the reactions made far less protein than reactions when plasmid was added prior to incubation (Figure 28). When reactions were incubated on ice for eight hours prior to plasmid addition, reactions using lysates ND and GD (which underwent dialysis) yielded higher GFP levels, though overall the GFP production was still less than the freshly assembled reactions, demonstrating that not even a significant decrease in temperature can stop the baseline metabolic activity of the CFE reaction components. From Figure 17, we see that dialysis, which is essentially a different type of long incubation of the lysate, actually helps the reaction. This supports the idea that the endogenous metabolism of the system is responsible for reaction degradation: dialysis is not done with all of the small molecule reaction supplements used in the incubation experiment in Figure 28, and thus it is the lysate's ability to act on these supplements that causes reaction degradation. This could be due to small molecule exhaustion or due to accumulation of toxic byproducts that are only produced when the small molecules are present in the reaction mixture.

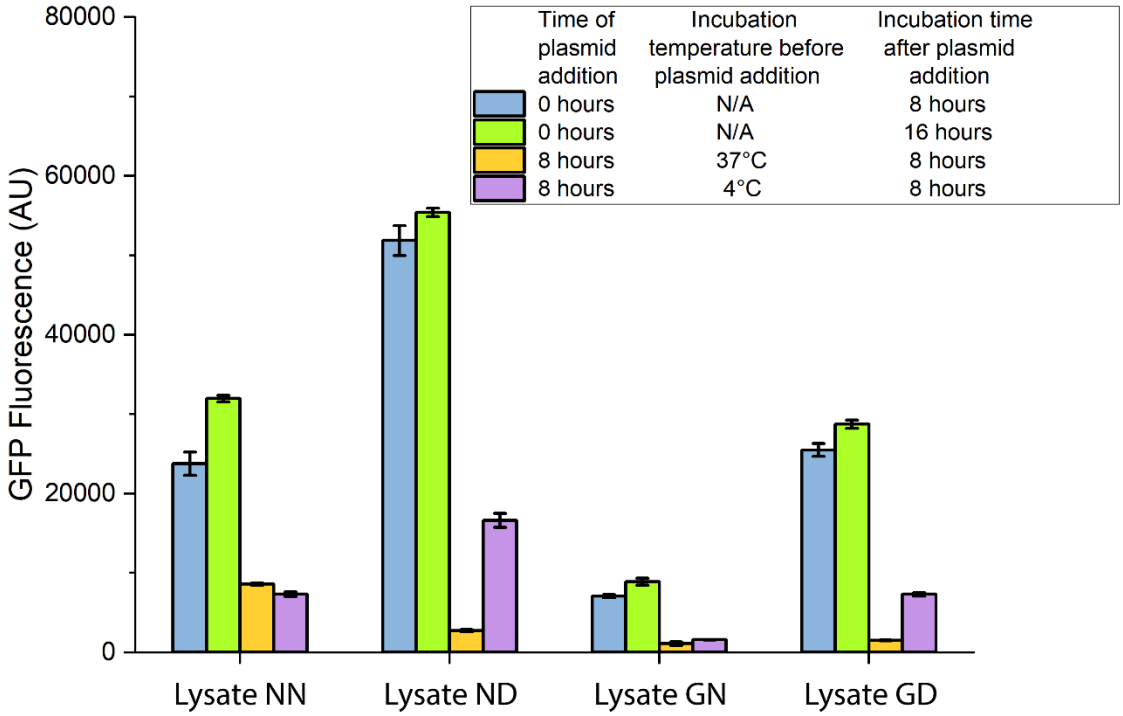


**Figure 28: GFP output of reactions run with plasmid supplementation at different times.**
**Preincubating CFE reactions for 8 hours at 37°C or 4°C prior to addition of pJL1s70 dramatically reduces protein expression. Error bars represent standard deviation of biological triplicates.**

### *3.3.4 Lysate preparation method and incubation time affect individual metabolites*

While an understanding of overarching metabolic changes could help to guide and standardize lysate processing and monitoring of CFE reactions, the insights from metabolomics analysis could also be applied to rational improvement of CFE reactions. Individual metabolites or combinations of metabolites that appear to be beneficial (or detrimental) to CFE performance could be added (or removed) from reactions to improve the output of CFE reactions. We aimed to identify potentially beneficial metabolites by analyzing how differentially accumulated individual metabolites correlate with lysate preparation and target protein production.

We first analyzed individual metabolite differences in the protein lysate via one-way ANOVA. As expected, the majority of metabolites were in a greater abundance in the non-dialyzed lysates (NN and GN) (Figure 29). However, a few metabolites (β-alanine, aspartate, and spermidine) were more concentrated in the dialyzed lysates, which is surprising, as these metabolites are not present in the dialysis buffer (Figure 30A-C). Since dialyzed lysates tend to produce more protein, these metabolites are of particular interest for designing improved reaction protocols, as supplementing metabolites to reactions is far easier than removing individual metabolites from prepared lysates. Thus, exploring potential molecular mechanisms for these metabolites could prove valuable.

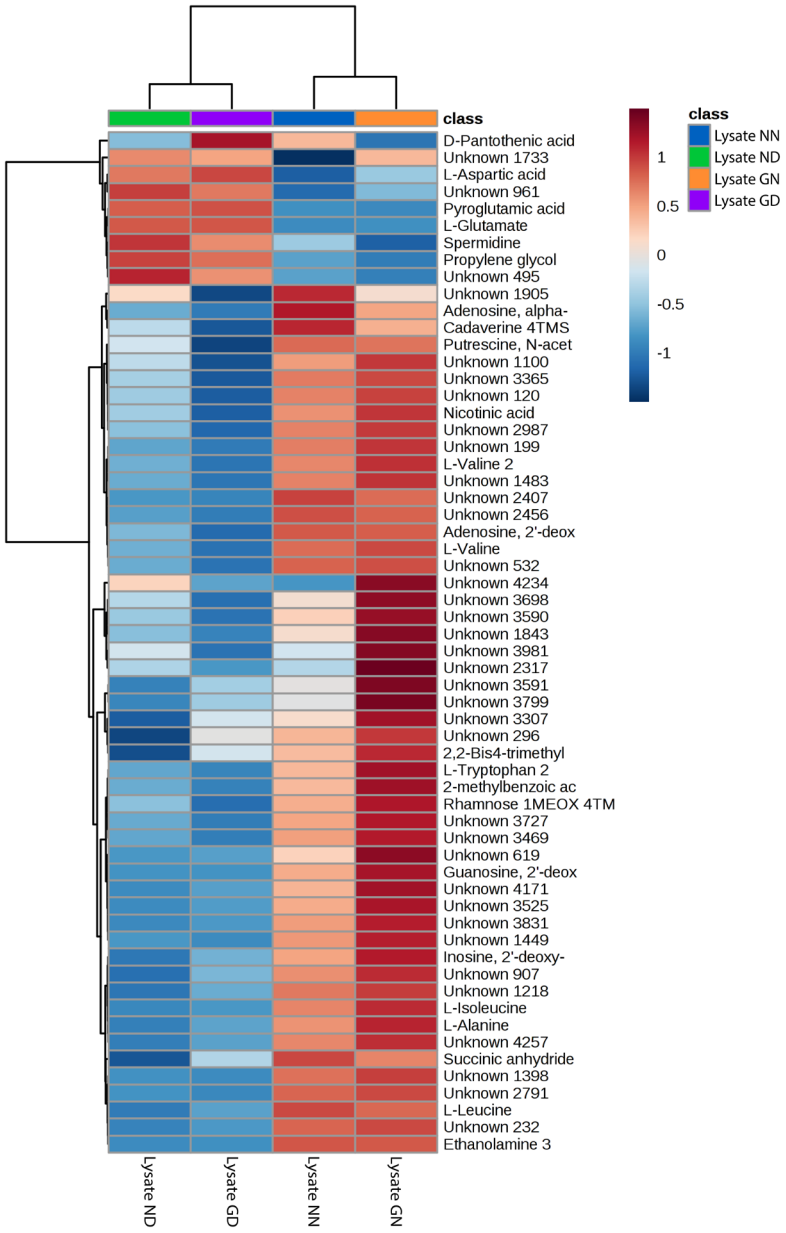


**Figure 29: Heatmap of metabolomics data collected from all lysates without reaction mixture or plasmid.**
**Clustering of lysates NN and GN from lysates ND and GD is driven by the effects of dialysis, with most metabolites depleted in the dialyzed samples. Euclidean distance and the Ward clustering algorithm were used for group clustering. Peak areas were log-transformed using generalized logarithm transformation (base 2) and autoscaled. Only the group averages of the top 60 analytes identified through ANOVA are displayed. Due to the high variability of β-alanine, it is not in the top 60 ANOVA-identified analytes.**

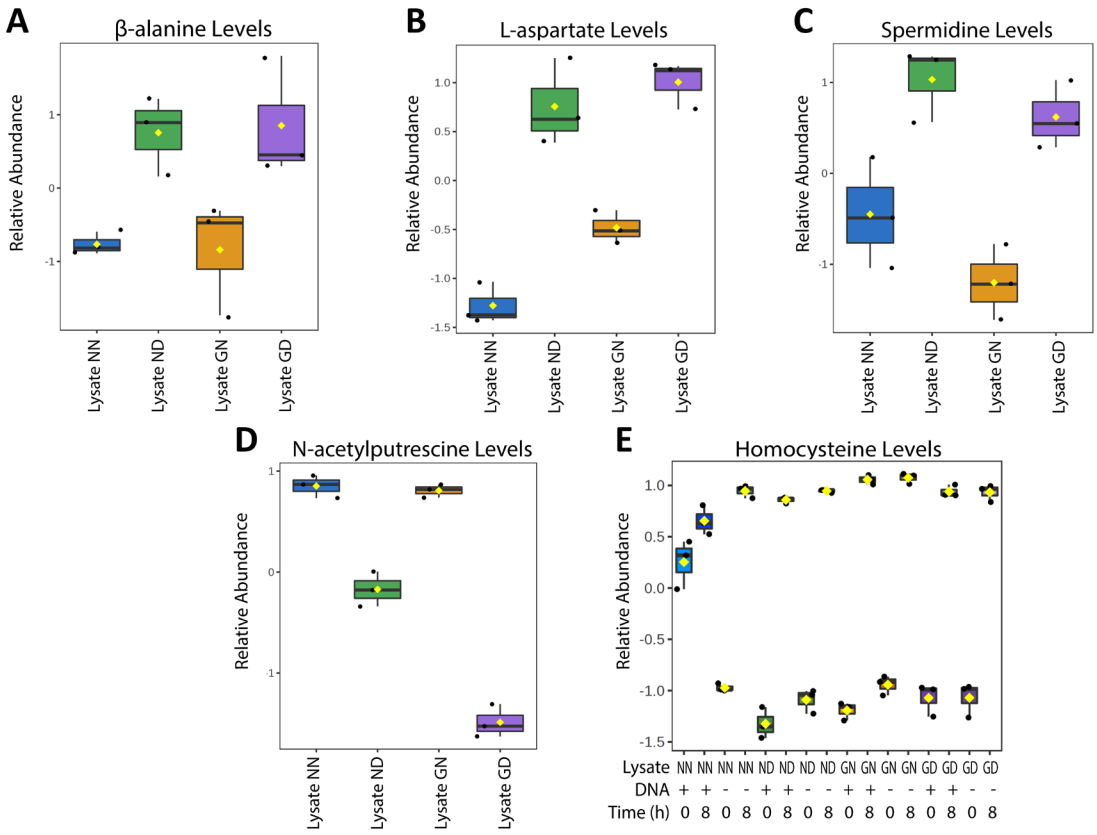


**Figure 30: Levels of individual metabolites vary in lysates and in completed CFE reactions.**
**Analysis of lysates (without reaction components) shows that (A) β-alanine, (B) L-aspartate, and (C) spermidine levels are all higher in dialyzed lysates (lysates ND and GD) than non-dialyzed lysates (lysates NN and GN), but (D) N-acetylputrescine levels are higher in non-dialyzed lysates. (E) Analysis of CFE reactions at different time points shows that homocysteine levels increase with time in all extracts, regardless of whether reactions produce GFP. Box and whisker plots depict the normalized peak areas, which are log-transformed using generalized logarithm transformation (base 2) and autoscaled. Black lines are the medians, boxes are the middle 50% values, black dots are individual sample levels, and yellow diamonds are the mean. Error bars represent standard deviation of biological triplicates.**

Two of these three metabolites— β-alanine and aspartate—are a part of the D-pantothenate (Vitamin B5) biosynthesis pathway, a precursor pathway for coenzyme A (CoA) production in *E. coli*[133]. CoA is a vital cofactor for a variety of pathways such as fatty acid biosynthesis and the TCA cycle[134] and is typically added to CFE reactions as part

of the reaction mixture. Though D-panthothenate levels were also measured in our experiments (Figure 31), β-alanine and aspartate are arguably more reliable markers for this pathway when using GC-MS due to known challenges in derivatizing D-pantothenate[135]. The third metabolite, spermidine, is a polyamine that interacts with RNA, promotes the assembly of the 30S ribosomal subunit, and stimulates protein synthesis[136]. Polyamines are also added as a part of the CFE reaction mixture, and previous studies have shown that further addition of polyamines can improve the productivity of *E. coli*-based CFE reactions[137, 138]. Interestingly, the relative abundances of putrescine, another polyamine, follow the opposite trend of spermidine: putrescine levels are lower in the dialyzed lysates, suggesting a potentially complex interplay of the effects of polyamines on CFE reaction output.

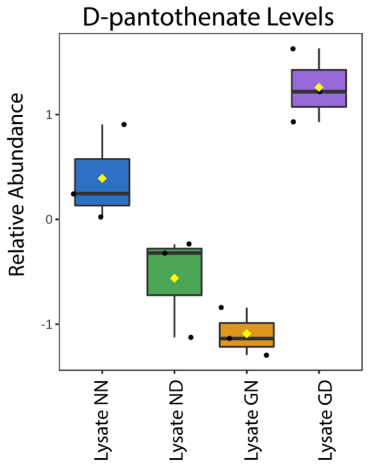


**Figure 31: Relative abundances of D-pantothenate in all lysates without reaction mixture or plasmid.**
**Box and whisker plots depict the normalized peak areas, which are log-transformed using generalized logarithm transformation (base 2) and autoscaled. Black lines are the medians, boxes are the middle 50% values, black dots are individual sample levels, and yellow diamonds are the mean. Error bars represent standard deviation of biological triplicates.**

Analysis of how individual metabolites change over the course of a CFE reaction (rather than just in the lysate used as part of the reaction) could further help to inform ways to extend the period or quantity of protein production. One-way ANOVA analysis revealed that homocysteine accumulates in all CFE reactions when protein production halts (Figure 30E). High levels of homocysteine are known to be toxic to whole *E. coli* cells[118, 139] via conversion to the intermediate homo-thiolactone that reduces protein functionality[140], which could also potentially inhibit transcription in CFE reactions. Pathway analysis, which reveals interconnectedness (in terms of metabolic pathway topology) of multiple differentially accumulated metabolites, also shows that two pathways—the CoA and pyruvate biosynthesis pathways—are significantly affected by time in all lysates (Table 7). Temporal changes in the group of metabolites related to CoA biosynthesis reinforces the findings of the individual CoA-related metabolite changes in dialyzed lysates as discussed above. Changes in pyruvate metabolism, which have previously been shown to occur in CFE reactions[66], are largely driven by the accumulation of lactate and propylene glycol, fermentative waste products that could indicate the depletion of glycolytic intermediates. Because glucose-6-phosphate (G6P) supplementation can alleviate glycolytic intermediate depletion-associated stress in cells[141], we hypothesized that G6P supplementation in CFE reactions could increase the availability of glycolytic intermediates and improve protein production.

**Table 7. Pathway analysis of metabolomics data collected from CFE reactions in all lysates.**

| Lysate Used in Reaction | Pathway Name | Match status | *p*-value | FDR |
|---|---|---|---|---|
| Lysate NN | Pyruvate metabolism | 6/26 | 3.3E-5 | 5.2E-4 |
| | Pantothenate and CoA biosynthesis | 5/23 | 0.0014 | 0.0056 |
| Lysate ND | Pyruvate metabolism | 6/26 | 8.6E-4 | 0.0038 |
| | Pantothenate and CoA biosynthesis | 5/23 | 0.011 | 0.028 |
| Lysate GN | Pyruvate metabolism | 6/26 | 7.7E-6 | 2.1E-5 |
| | Pantothenate and CoA biosynthesis | 4/23 | 3.0E-5 | 6.1E-5 |
| Lysate GD | Pyruvate metabolism | 6/26 | 2.2E-4 | 0.0018 |
| | Pantothenate and CoA biosynthesis | 5/23 | 0.0028 | 0.0083 |

### *3.3.5 Supplementation of identified metabolites change affects lysate activity*

With a list of potentially critical metabolites in hand, we next tested whether supplementation of these metabolites could improve the performance of CFE reactions. We specifically tested the ways that spermidine, putrescine, β-alanine, homocysteine, and G6P affected CFE reactions by adding them to reactions that contain pJL1s70 and then quantifying GFP production at 8 hours. We added concentrations of β-alanine, homocysteine, and G6P[141] based on reported intracellular concentrations. Since putrescine and spermidine are already added to CFE reactions at a concentration of 1.0 and 1.5 mM, respectively, we added higher concentrations of these compounds.

Of all components tested, putrescine elicited the largest improvement in protein yields in all lysates (Figure 32A). Spermidine, another polyamine, only improved CFE output when supplemented at 5 mM to reactions containing lysates NN and ND and nearly completely inhibited protein synthesis at 20 mM, emphasizing the importance of optimizing the concentrations of supplemented molecules. β-alanine supplementation modestly improved the performance of reactions with lysate NN and ND, suggesting that its role could perhaps be related to the composition of the growth medium. Homocysteine supplementation also altered the output of CFE reactions, but in an unexpected way: counter to our initial hypothesis that homocysteine decreases CFE performance, homocysteine supplementation improved performance of all reactions. G6P addition had little to no effect on reaction performance. Interestingly, addition of putrescine to reactions with lysate NN and GD enabled them to produce higher protein than unsupplemented reactions with lysate ND (the lysate that initially had the greatest GFP production), suggesting that putrescine could be titrated into reactions to adjust lysate activity and compensate for differences in preparation (Figure 32B).

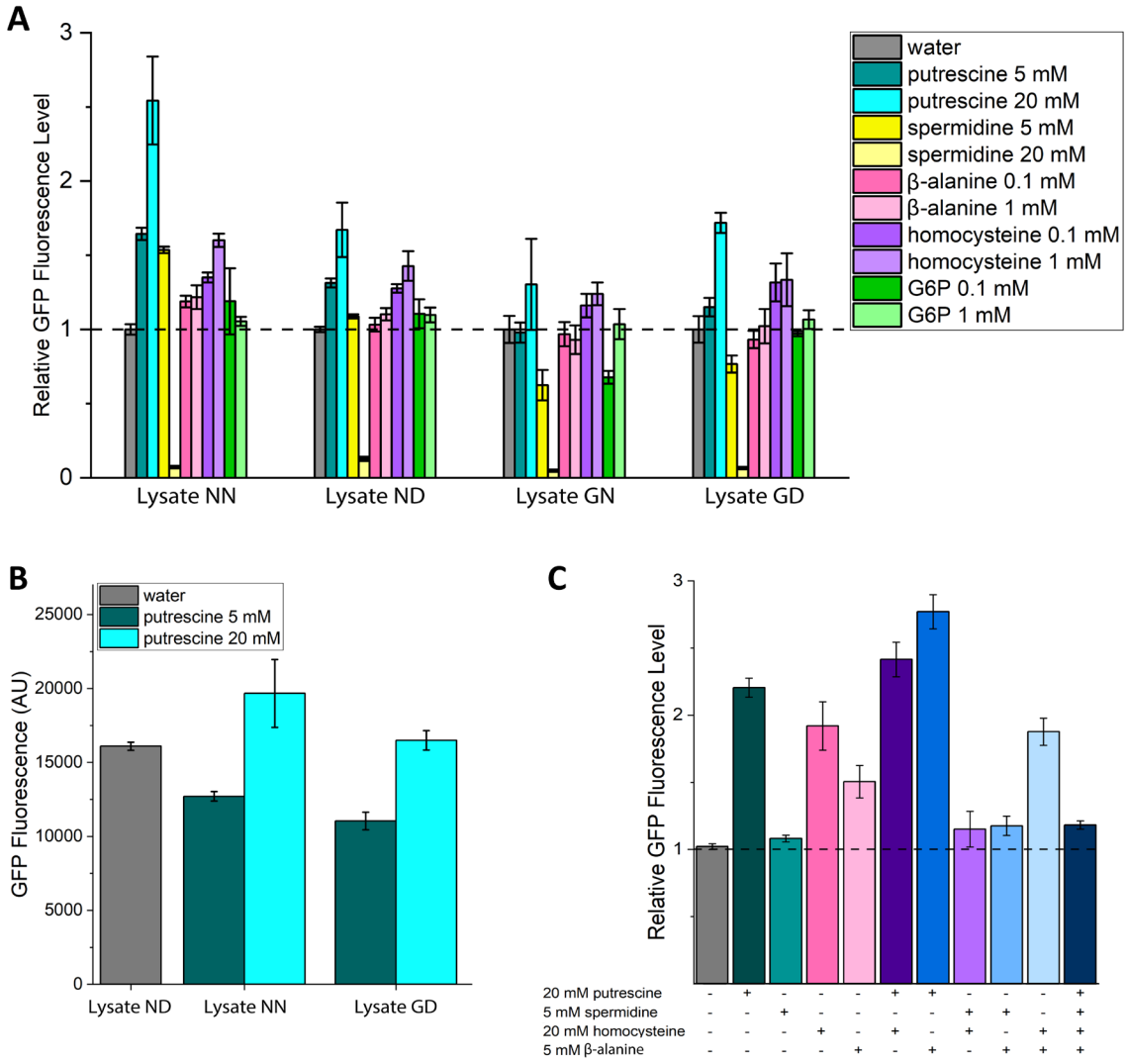


**Figure 32: Metabolite supplementation of CFE reactions can strongly impact GFP production.**

**(A) The metabolites putrescine, spermidine, β-alanine, homocysteine, and G6P were added to CFE reactions. Fluorescence relative to reactions run in the same lysate without additional components was calculated. Addition of putrescine to CFE reactions most strongly improved GFP expression. (B) Putrescine supplementation can normalize output across CFE reactions that contain differentially-prepared lysates. Addition of 20 mM putrescine to reactions with either lysate NN and lysate GD cause them to produce more GFP than reactions with lysate ND and no added metabolites. (C) Combinatorial metabolite supplementation further improves CFE performance. Addition of putrescine with either β-alanine or homocysteine improved CFE output relative to addition of either individual component. Error bars represent standard deviation of technical triplicates.**

With the goal of creating a CFE system with maximal activity, we next focused on optimizing the output of reactions run with lysate ND. We first tested whether higher concentrations of β-alanine and homocysteine could further improve CFE performance and found that 5 mM β-alanine and 20 mM homocysteine most strongly increased protein expression (Figure 33). Finally, we tested pair-wise combinations of putrescine, spermidine, β-alanine, and homocysteine. Addition of putrescine in combination with either β-alanine or homocysteine further improved lysate performance, enabling us to create a CFE system that produces over 2.5-fold more protein than the original system (Figure 32C). We further note that the effects of the addition of individual molecules are not additive (Table 8, Table 9, & Table 10), indicating the complexity of interactions in CFE reactions and the difficulty that future optimization of these reactions will entail.

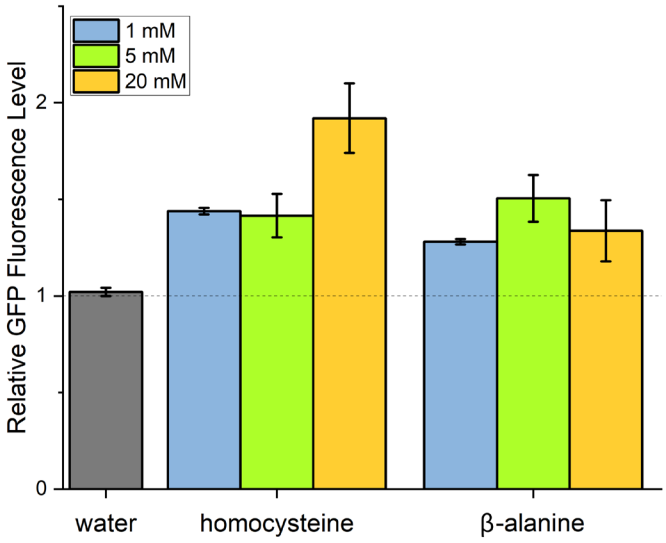


**Figure 33: Optimization of homocysteine and β-alanine supplementation levels in lysate ND.**
**Different concentrations of homocysteine and β-alanine were added to CFE reactions run in lysate ND, and fluorescence was measured at 8 hours. 20 mM homocysteine and 5 mM β-alanine most strongly improve GFP production. Error bars represent standard deviation of technical triplicates.**

**Table 8: Description of the general equation to determine the interaction effects of supplemented metabolites.**

| **General Equation:** $Y = \mu_0 + Ax_{1_i} + Bx_{2_i} + Cx_{3_i} + Dx_{4_i} + I_{12}x_{1_i}x_{2_i} + I_{13}x_{1_i}x_{3_i} + I_{14}x_{1_i}x_{4_i} + I_{23}x_{2_i}x_{3_i} + I_{24}x_{2_i}x_{4_i} + I_{34}x_{3_i}x_{4_i} + I_{all}x_{1_i}x_{2_i}x_{3_i}x_{4_i}$ | |
|---|---|
| **Variable** | **Description** |
| $Y$ | GFP Output |
| $\mu_0$ | constant |
| $A$ | 1st-order effect coefficient for 20 mM putrescine |
| $B$ | 1st-order effect coefficient for 5 mM spermidine |
| $C$ | 1st-order effect coefficient for 20 mM homocysteine |
| $D$ | 1st-order effect coefficient for 5 mM β-alanine |
| $I_{12}$ | 2nd-order interaction effect coefficient for 20 mM putrescine and 5 mM spermidine |
| $I_{13}$ | 2nd-order interaction effect coefficient for 20 mM putrescine and 20 mM homocysteine |
| $I_{14}$ | 2nd-order interaction effect coefficient for 20 mM putrescine and 5 mM β-alanine |
| $I_{23}$ | 2nd-order interaction effect coefficient for 5 mM spermidine and 20 mM homocysteine |
| $I_{24}$ | 2nd-order interaction effect coefficient for 5 mM spermidine and 5 mM β-alanine |
| $I_{34}$ | 2nd-order interaction effect coefficient for 20 mM homocysteine and 5 mM β-alanine |
| $I_{all}$ | 4th-order interaction effect coefficient for 20 mM putrescine, 5 mM spermidine, 20 mM homocysteine, and 5 mM β-alanine |
| $x_{1_i}$ | Represents the presence (+1) or absence (-1) of 20 mM putrescine in reaction i |
| $x_{2_i}$ | Represents the presence (+1) or absence (-1) of 5 mM spermidine in reaction i |
| $x_{3_i}$ | Represents the presence (+1) or absence (-1) of 20 mM homocysteine in reaction i |
| $x_{4_i}$ | Represents the presence (+1) or absence (-1) of 5 mM β-alanine in reaction i |

**Table 9: Calculated interaction effects of supplemented metabolites.**

| Variable | Coefficients | t Stat | *p*-value | Lower 95% | Upper 95% |
|---|---|---|---|---|---|
| $\mu_0$ | 39700.94 | 81.39 | 7.90E-31 | 38694.22 | 40707.66 |
| $A$ | 7009.88 | 14.37 | 2.75E-13 | 6003.16 | 8016.59 |
| $B$ | -9662.29 | -19.81 | 2.23E-16 | -10669.01 | -8655.57 |

| | | | | | |
|---|---|---|---|---|---|
| $C$ | 1297.96 | 2.66 | 1.37E-02 | 291.24 | 2304.68 |
| $D$ | 3551.96 | 7.28 | 1.60E-07 | 2545.24 | 4558.68 |
| $I_{12}$ | -4627.52 | -9.49 | 1.36E-09 | -5634.24 | -3620.80 |
| $I_{13}$ | -2890.94 | -5.93 | 4.08E-06 | -3897.66 | -1884.22 |
| $I_{14}$ | 1678.23 | 3.44 | 2.13E-03 | 671.51 | 2684.95 |
| $I_{23}$ | -3581.60 | -7.34 | 1.39E-07 | -4588.32 | -2574.89 |
| $I_{24}$ | -972.10 | -1.99 | 5.78E-02 | -1978.82 | 34.61 |
| $I_{34}$ | -1701.27 | -3.49 | 1.90E-03 | -2707.99 | -694.55 |
| $I_{all}$ | -1198.23 | -2.46 | 2.16E-02 | -2204.95 | -191.51 |

**Table 10: Calculated interaction effects of supplemented metabolites ignoring 4$^{th}$-order interaction effects.**

| Variable | Coefficients | t Stat | *p*-value | Lower 95% | Upper 95% |
|---|---|---|---|---|---|
| $\mu_0$ | 36106.25 | 26.89 | 2.53E-18 | 33321.89 | 38890.61 |
| $A$ | 4613.42 | 4.55 | 1.59E-04 | 2508.64 | 6718.19 |
| $B$ | -12058.75 | -11.88 | 4.81E-11 | -14163.52 | -9953.98 |
| $C$ | -1098.50 | -1.08 | 2.91E-01 | -3203.27 | 1006.27 |
| $D$ | 1155.50 | 1.14 | 2.67E-01 | -949.27 | 3260.27 |
| $I_{12}$ | -5825.75 | -8.12 | 4.63E-08 | -7314.05 | -4337.45 |
| $I_{13}$ | -4089.17 | -5.70 | 9.90E-06 | -5577.47 | -2600.87 |
| $I_{14}$ | 480.00 | 0.67 | 5.11E-01 | -1008.30 | 1968.30 |
| $I_{23}$ | -4779.83 | -6.66 | 1.07E-06 | -6268.13 | -3291.53 |
| $I_{24}$ | -2170.33 | -3.02 | 6.23E-03 | -3658.63 | -682.03 |

| | | | | | |
|---|---|---|---|---|---|
| $I_{34}$ | -2899.50 | -4.04 | 5.47E-04 | -4387.80 | -1411.20 |

### 3.4 Limitations

While GC-MS-based metabolomics analysis is quite valuable, it is inherently limited to detection of molecules that are either volatile or volatilizable by standard derivatization reagents. Also, GC-MS cannot accurately capture di-, bis-, or triphosphates such as NADH, ATP, and other cofactors. Understanding how these types of molecules are consumed would greatly inform which molecules are potentially limiting CFE reactions and better inform supplementation efforts. Other analysis methods such as liquid chromatography-mass spectrometry could complement GC-MS analyses by capturing nonvolatile and nonderivatizable metabolites as well as cofactors. Heavy isotope-labeled standards could also be used to obtain quantitative metabolite concentrations (rather than relative abundances), which would both lend more precision to analyses and also better inform concentrations of metabolites to supplement. Proteomics could complement metabolomics analysis on the same samples by revealing whether specific enzymes are present at different levels in crude lysates[64, 65] prepared via different protocols, which could then be related to the enzymatic activity captured by metabolomics.

Similarly, we supplemented only identified metabolites with obvious potential biological significance at concentrations based on broad assumptions and previously reported intracellular levels. Identification and subsequent testing of unannotated analytes in conjunction with testing larger concentration ranges may lead to the creation of further improved CFE systems. The small scale and “open system” nature of CFE reactions make high-throughput assessment of additives feasible, and tools that enable automated

metabolic analysis of samples in real-time could enable a temporal systems-wide analysis of the effect of supplementation. Moreover, the large data sets generated through such system-wide analyses and high-throughput screens could be combined with computational models and machine learning to better inform standardization and optimization of CFE systems for a variety of applications, including biosensing.

## 3.5 Conclusion

In summary, we used metabolomics via GCxGC-MS analysis to identify difference in the levels of small molecules present in differently prepared lysates, to understand their potential roles in lysate activity, and to inform lysate optimization. Through a systems-level analysis of metabolites, we confirmed the expected result that dialysis most strongly affects the metabolite profile of isolated protein lysates. To our surprise, we also found that the endogenous metabolic activity of the lysate overwhelms the metabolic changes caused by protein synthesis in CFE reactions. We then explored how changes to the initial metabolic state of the lysate affect CFE performance by incubating reactions without plasmid DNA and found that the protein production capacity of the incubated reactions was majorly reduced compared to that of freshly assembled reactions, demonstrating that the endogenous metabolism of the system is responsible for reaction degradation. In-depth analysis of individual metabolite changes revealed that the pantothenate precursors (β-alanine and aspartate) and polyamines (putrescine and spermidine) differentially accumulated based on lysate preparation method and that levels of homocysteine and metabolites involved in pyruvate and pantothenate metabolism changed consistently by the end of protein synthesis for all reactions, regardless of the type of lysate used. Based on this information, we sought to understand the roles of these molecules in CFE activity via

metabolite supplementation. We found that individual supplementation of putrescine and homocysteine resulted in significant improvements to protein production, and that the addition of putrescine in combination with either β-alanine or homocysteine further improved lysate performance, increasing CFE output by a factor of 2.5 relative to the initial preparation protocol. Our results highlight the importance of small molecules on CFE productivity and underscore the importance of fully understanding and controlling endogenous metabolic activity to improve CFE reactions and biosensor development.

# CHAPTER 4. METABOLIC DYNAMICS OF *ESCHERICHIA COLI*-BASED CELL-FREE EXPRESSION SYSTEMS

## 4.1 Introduction

In the previous chapter, we used metabolomics to assess the impacts of differently prepared *E. coli*-based CFE lysates and found that the reactions composed of these different lysates were metabolically distinct[43]. To our surprise, we found that the metabolite-level changes at the end of a CFE reaction due to endogenous metabolic activity in the reactions eclipsed any metabolic changes due to protein synthesis. These findings highlight the complexity of CFE systems, our lack of understanding of their metabolic underpinnings, and the resultant need for broader metabolic investigations of CFE systems to ultimately enabled improved biosensor development.

Although we were able to identify a handful of small molecules that impacted reaction activity and demonstrated that certain metabolites could be supplemented into the lysates to improve reaction performance, it is still unclear how much the lysate's endogenous enzymes and metabolites impact the productivity of a CFE reaction. One prominent hypothesis is that CFE reaction efficiency and lifetime is significantly (or even predominantly) affected by the depletion or accumulation of specific metabolites that affect protein synthesis[39, 50], though the identities of all the molecules involved is not known. While it is known that many central carbon metabolism enzymes are present in *E. coli*-based cell-free lysates[64-66], it is largely unknown how they impact metabolic processes downstream of central carbon metabolism and alter the transcriptional and translational

capacity of a CFE reaction (Figure 34). Unfortunately, unlike in living whole-cell biological systems, we have minimal understanding of the endogenous metabolism in CFE systems, prompting a serious need for a more thorough characterization of their metabolic dynamics. Deeper understanding of CFE metabolism would facilitate not only biosensor optimization efforts, but also rational approaches to resolve reproducibility, scalability, and standardization issues in CFE systems.

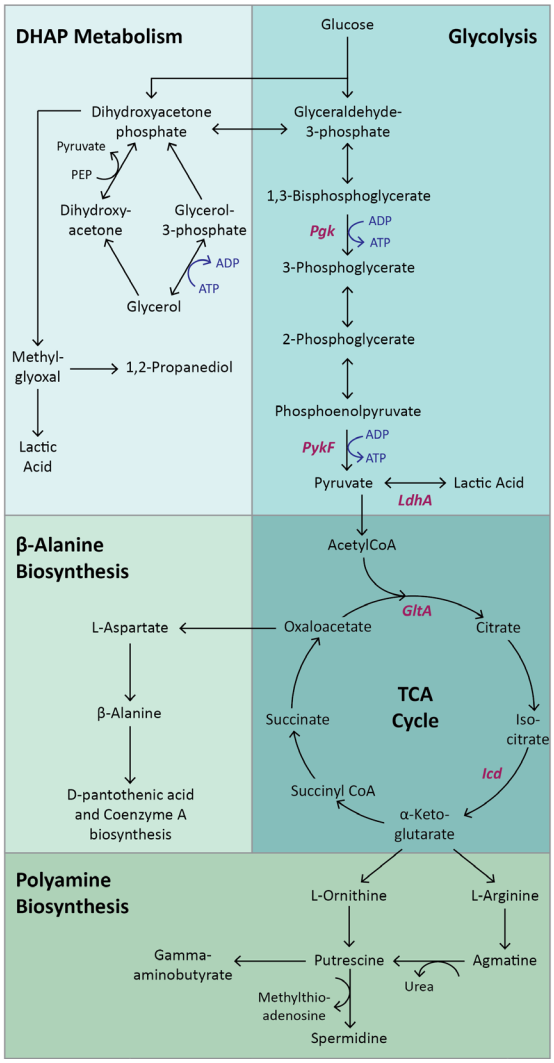


**Figure 34: An overview of metabolic pathways relevant to this work.**
**Enzymes of interest specifically discussed in this paper are noted in italics. Metabolic pathways and metabolites are as follows: DHAP, dihydroxyacetone phosphate; TCA,**

**tricarboxylic acid cycle; ATP, adenosine triphosphate; ADP, adenosine diphosphate; PEP, phosphoenolpyruvate; The enzymes shown are as follows Pgk, phosphoglycerate kinase; PykF, pyruvate kinase; LdhA, lactate dehydrogenase; GltA, citrate synthase; Icd, isocitrate dehydrogenase.**

Here, we use metabolomics to more broadly characterize the metabolic profiles of *E. coli*-based CFE systems. In particular, we aim for a deeper characterization of the metabolic dynamics of these systems, as previous analysis focused on long end-point times and thus may have missed critical dynamics during protein synthesis and as CFE activity declined. We also deconstruct the CFE reaction into its constituent components for metabolomic analysis, in an effort to more clearly pinpoint the source of metabolic changes observed in complete CFE reactions. We explore the effects of lysate pre-incubation and how changes to the sonication energy input during lysate preparation affect both endogenous metabolism and protein synthesis in CFE reactions. Finally, we use the information from these studies to select native metabolic enzymes to supplement in CFE reactions in an effort to alter metabolic activity and thus protein yield of a reaction.

## 4.2 Experimental Methods

### *4.2.1 Plasmids*

The plasmids pJL1s70 and E01 were used in this study. Plasmids were transformed into *E. coli* DH10B cells and isolated with E.Z.N.A FastFilter Plasmid Maxiprep kit (Omega Biotek) according to the manufacturer's instructions.

### *4.2.2 CFE lysate preparation*

Cellular lysate for all experiments was prepared based on previously described protocols[39, 41]. Briefly, BL21 cells were used for all experiments, except for the experiment

evaluating the effects of sonication energy input, where BL21 DE3 Star Δ*lacZ* cells were used. All cells were grown in either 2x YTP media (16 g $L^{-1}$ tryptone, 10 g $L^{-1}$ yeast extract, 5 g $L^{-1}$ sodium chloride, 7 g $L^{-1}$ potassium phosphate dibasic, and 3 g $L^{-1}$ potassium phosphate monobasic and was pH-corrected to 7.2 with Tris base). All media was filter-sterilized prior to use. Cells were grown at 37 °C and 180 rpm to an OD of 2.0, which corresponds with the mid exponential growth phase. Cells were then centrifuged at 2700 rcf and washed three times with S30A buffer (14 mM magnesium acetate, 60 mM potassium acetate, 10 mM Tris-acetate (pH 8.2), and 2 mM dithiothreitol). After the final centrifugation, the wet cell mass was determined, and cells were resuspended in 1 mL of S30A buffer per 1 g of wet cell mass. The cellular resuspension was divided into 1 mL aliquots. Cells were lysed using a Q125 Sonicator (Qsonica, Newton, CT) at a frequency of 20 kHz and at 50% of amplitude. Cells were sonicated on ice with three cycles of 10 s on, 10 s off, delivering approximately 300 J unless otherwise specified in text. An additional 4 mM of dithiothreitol was added to each tube, and the sonicated mixture was then centrifuged at 12 000 rcf and 4 °C for 10 min. For lysates prepared with BL21 cells, the supernatant was removed and divided into 1 mL aliquots for run-off reaction at 37 °C and 180 rpm for 80 min. After this runoff reaction, the cellular lysate was centrifuged at 12,000 rcf and 4 °C for 10 min. The supernatant was removed and loaded into a 10 kDa MWCO dialysis cassette (Thermo Scientific). Lysate was dialyzed in 1L of S30B buffer (14 mM magnesium glutamate, 60 mM potassium glutamate, 1 mM dithiothreitol, pH-corrected to 8.2 with Tris) at 4 °C for 3 hours. Dialyzed lysate was removed and centrifuged at 12,000 rcf and 4 °C for 10 min. The supernatant was removed, aliquoted, and stored at −80 °C for future use. For lysates prepared with BL21 DE3 Star Δ*lacZ* cells, the

supernatant was removed and divided into 0.5 mL aliquots for run-off reaction at 37 °C and 220 rpm for 80 min. All downstream processing steps were the same as those for lysates prepared with BL21 cells, with the exception of 500 mL of S30B buffer used during dialysis.

### *4.2.3 Protein purification*

Plasmids coding for expression of different his-tagged proteins were transformed into BL21 (DE3) cells and plated on LB plates supplemented with 33 μg/mL kanamycin to grow overnight. One colony was selected the next day and resuspended in a 50 mL LB culture supplemented with 33 μg/mL kanamycin for overnight growth. The overnight culture was then diluted 100-fold in 500 mL of fresh 2xYTP media containing kanamycin the next morning and grown until its $OD_{600}$ reached between 0.4-0.6, at which point 0.4 mM of IPTG was added to induce T7 polymerase expression and thus plasmid-driven protein production. The induced culture was transferred to a shaking water bath for incubation at 30 ˚C and 180 rpm for 16 hours, when cells were pelleted, weighed, and frozen at -80 ˚C for storage until cell lysis.

1 g of frozen cell pellet was resuspended in 2 mL of lysis buffer (50 mM $Na_2HPO_4$, 500 mM NaCl, 10 mM imidazole, pH 8). The resuspension was divided into 1 mL aliquots and sonicated until cells appeared visible lysed. Sonicated products were centrifuged at 12,000 rcf and 4 ˚C for 15 min before purifying on a HisPur Ni-NTA column (Thermo Scientific) according to the manufacturer protocol. Purification was verified by SDS-PAGE. The eluted proteins were loaded into 10 kDa MWCO dialysis cassettes (Thermo Scientific) and dialyzed overnight in the storage buffer (50 mM Tris-HCl pH 7.5, 100 mM

NaCl, 1 mM DTT, 1 mM EDTA, 2% DMSO). Following dialysis, proteins were centrifuged at 12,000 rcf at 4 ˚C for 10 min to remove insoluble fractions. The supernatant was removed, and protein concentration was measured on a Nanodrop 2000 before sub-aliquoting and storage at -20 ˚C.

### *4.2.4 CFE reactions and assessment of GFP production*

Cell-free reactions for all experiments were run as previously described[41]. Each cell-free reaction contained 0.85 mM each of GTP, UTP, and CTP, in addition to 1.2 mM ATP, 34 μg/mL of folinic acid, 170 μg/mL E. coli tRNA mixture, 130 mM potassium glutamate, 10 mM ammonium glutamate, 12 mM magnesium glutamate, 2 mM each of the 20 standard amino acids, 0.33 mM nicotine adenine dinucleotide (NAD), 0.27 mM coenzyme-A (CoA), 1.5 mM spermidine, 1 mM putrescine, 4 mM sodium oxalate, 33 mM phosphoenolpyruvate (PEP), 27% cell lysate, and 12 nM of the specified plasmid. (9 nM of pJL1s70 was used for the experiment evaluating the effects of sonication energy input.)

For metabolomics analysis, 210 μL reactions were prepared in 1.5 mL microcentrifuge tubes in technical triplicates. Samples were incubated at 37 °C for the specified time. A total of 10 μL of the reaction was then removed and stored at −80 °C for subsequent fluorescence analysis on a BioTek Synergy H4 microplate reader (485/510 nm excitation/emission wavelength, gain of 70). and the remaining 200 μL was stored at −80 °C for subsequent metabolomics analysis. In experiments solely assessing GFP production, 10 μL reactions were prepared in technical triplicates in 384-well plates (Greiner Bio-One) and fluorescence values were measured every 5 minutes at 37 ˚C. A transparent film was used to seal the plates to prevent reagent evaporation.

### *4.2.5 Protein precipitation for metabolomics analysis*

Before beginning the protein precipitation protocol, a small volume was removed from all samples in an individual experiment to prepare pooled quality control (QC) samples for the mass spectrometry data acquisition. 25 µL was removed from each sample from the in-depth time course analysis of CFE reactions and the comparison of reactions with differently sonicated lysates. 20 µL was removed from each sample for the time course analysis of the incubated lysate and the comparison of enzyme-supplemented reactions. 15.4 and 10 µL were removed from each samples for the time course analysis of the incubated reaction mix and the comparison of reactions with lysates pre-incubated at different temperatures, respectively. These pooled QC samples were prepared with all other samples for protein precipitation.

Proteins were precipitated from all samples stored for metabolomics analysis via the following protocol[74]: first, methanol was added to each sample at a 1:2 sample to methanol ratio and vortexed briefly. The samples were incubated at −20 °C for 20 min and centrifuged at 11,600 rcf for 30 min at room temperature, and the supernatant was collected. The supernatants of pooled QC samples were then evenly aliquoted into multiple tubes as needed: two tubes for the in-depth time course analysis of CFE reactions and the time course analysis of the incubated lysate; three tubes for the comparison of reactions with lysates pre-incubated at different temperatures, the comparison of enzyme-supplemented reactions, and the comparison of reactions with differently sonicated lysates; and one tube for the time course analysis of the incubated reaction mix. The supernatants were dried at 40 °C in a CentriVap until all water was removed and stored at −80 °C.

### *4.2.6 GC-MS analysis*

Before derivatization, stored samples were transferred to a CentriVap to be dried at 40 °C for 15 min. Samples were derivatized as previously described[99, 128]. A total of 10 µL of 40 mg/mL *O*-methylhydroxylamine hydrochloride (MP Biomedicals, LLC, Santa Ana, CA, U.S.A.) in pyridine was added to each dried sample and shaken at 1400 rpm for 90 min at 30 °C. A total of 90 µL of N-methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) + 1% trimethylchlorosilane (TMCS) (Thermo Scientific, Lafayette, CO, U.S.A.) was then added to the samples and shaken at 1400 rpm for 30 min at 37 °C. Samples were centrifuged at 21,100 rcf for 3 min, and 50 µL of the supernatant was added to an autosampler vial. Samples were spiked with 0.25 µL of a retention time standard solution composed of fatty acid methyl esters (FAMES). At the beginning of the GC-MS run, the QCs were injected once, and this was repeated again after every 4−6 sample injections to allow for downstream correction for batch effects. A derivatization blank was prepared and run with every batch of samples. A LECO Pegasus 4D instrument with an Agilent 7683B autosampler, Agilent 7890A gas chromatograph, and time-of-flight mass spectrometer (TOF-MS) was used to analyze the samples. The first column was an HP-5, 28 m long × 0.320 mm ID × 0.25 µm film thickness (Agilent, Santa Clara, CA, U.S.A.), and the second was an Rtx-200, 1.5 - 1.8 m long × 0.25 mm ID × 0.25 µm film thickness (Restek, Bellefonte, PA, U.S.A.). More detailed gas chromatography, autosampler, and mass spectrometry methods are provided in Appendix A.

### *4.2.7 Data analysis*

Sample runs were analyzed in ChromaTOF (LECO, St. Joseph, MI, U.S.A.) to determine baseline, peak area, and peak identification as described previously[101, 102]. Briefly, settings included a baseline offset of 0.5, automatic smoothing, first dimension peak width of 36 s, second dimension peak width of 0.10 s, and a match of 700 required to combine peaks with a minimum signal-to-noise (S/N) of 5 for all subpeaks. Peaks were required to have a S/N of 10 and have a minimum similarity score of 800 to NIST, Golm, and in-house spectral libraries before assigning a name. Unique mass was used for area and height calculation. MetPP was used to align the samples[103]. Sample files and a derivatization reagent blank file were uploaded from ChromaTOF. Unknowns were retained during the peak alignment process. The derivatization reagent blank file was used to subtract peaks resulting from the sample preparation reagents from the corresponding sample files. On-the-fly alignment was used with manually selected quality control samples as the peak list for primary alignment. Peak alignment was performed using the default criteria. To remove analytes that were not reproducibly detected, analytes for which more than half of the values were missing in the QC samples or for which the QC samples had a coefficient of variance larger than 0.5 were removed from the data set. Then, missing values were manually corrected using small value correction only if all the values were missing in the biological replicates.

Finally, MetaboAnalyst was used for statistical and two-factor analysis[142]. For both analyses, remaining missing values were k- nearest neighbors (KNN) corrected. Data was then log-transformed using a generalized logarithm (base 2) and autoscaled. *P*-values were adjusted using the Benjamini-Hochberg False Discovery Rate (FDR). Differences were considered significant at FDR-corrected p-values $< 0.05$.

The metabolomics data sets for this study are available via the Metabolights repository, with the data set identifier MTBLS2630[143].

## 4.3 Results and Discussion

### *4.3.1 Metabolic profiles change throughout CFE reactions*

Our first goal was to characterize the metabolic dynamics in a CFE reaction for the duration of protein synthesis, as previous studies had focused on reaction endpoints. We prepared the lysate from exponentially growing BL21 cells in 2xYTP media; the cells were lysed via sonication, and the lysate was post-processed with a run-off reaction and dialysis (see Experimental Methods for details). A CFE reaction was assembled comprising this lysate, a small-molecule reaction mixture, and the CFE plasmid pJL1s70 to drive expression of green fluorescent protein (GFP) from an *E. coli* $\sigma^{70}$ promoter. GFP production was measured and samples were collected for metabolomics analysis at 0, 0.5, 1, 2, 4 and 6 hours. Metabolomics samples were prepared by precipitating proteins and analyzing the remaining metabolite mixture using two-dimensional gas chromatography coupled to mass spectrometry (GCxGC-MS) after sample derivatization. The resulting instrument output was processed with a computational workflow resulting in relative abundances for 276 putatively identified and unannotated metabolites that were used for downstream analysis.

For a systems-scale analysis of the temporal metabolic changes in the reaction, we analyzed the resulting data with principal component analysis (PCA). For the time-course CFE profiles, the majority of sample group separation is captured in the first principal component (PC1), reflecting a monotonic change in metabolic state over the course of the

reaction (Figure 35B). The fact that the 0 hour and 0.5 hour samples are the most separated consecutive timepoints in PC1 suggests that a large portion of the metabolic changes likely occurred in the first half hour of the reaction; however, the separation of later timepoint samples in this same principal component suggests that metabolite levels continue to change throughout the entirety of the reaction. Surprisingly, metabolite profiles at 4 hours separate from those at 6 hours in both PC1 and PC2, showcasing that significant metabolic activity continues even as protein synthesis is concluding **(**Figure 35A).

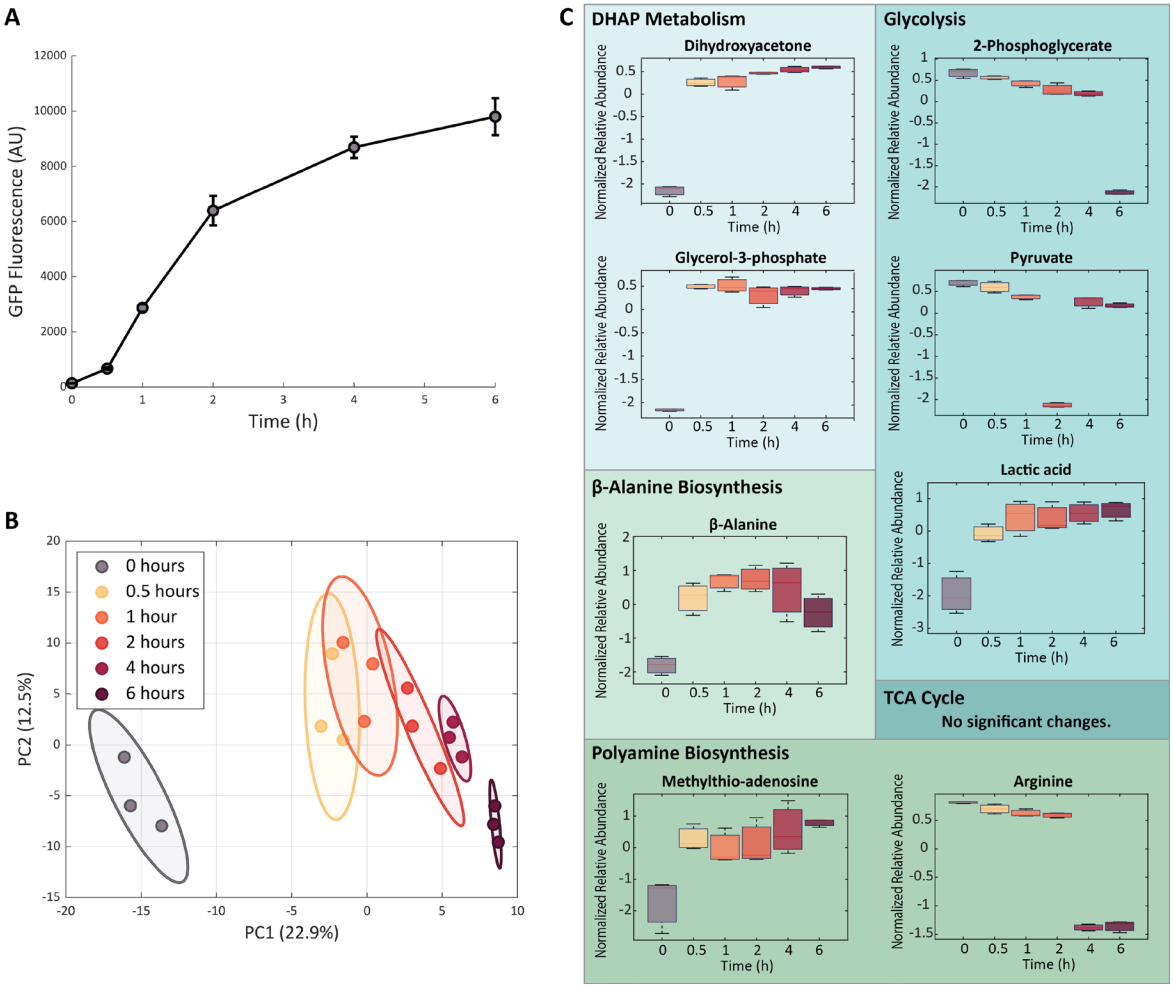


**Figure 35: Temporal profiles of CFE reactions during protein production. (A) GFP production (measured via fluorescence) slows down at around 4 hours. Error bars represent standard deviation of triplicate reactions. (B) Samples from different**

**timepoints separate from one another, with PC1 values increasing with reaction time. Colored ellipses represented 95% confidence intervals for each group, and the plotted samples are triplicate reactions. (C) Profiles of metabolites involved in glycolysis, dihydroxyacetone phosphate (DHAP) metabolism, β-alanine biosynthesis, and polyamine biosynthesis pathways change over the course of a CFE reaction. Box and whisker plots depict the normalized peak areas, which are transformed using a generalized logarithm (base 2) and autoscaled. Red lines are the medians, boxes span the second and third quartiles of values. Error bars represent standard deviation of triplicate reactions.**

### *4.3.2 Levels of multiple key metabolic pathways evolve over the course of a CFE reaction*

We then identified individual metabolites with significant changes during the reaction using ANOVA. We found major metabolic changes in central carbon and amino acid metabolism. Specifically, we detected significant changes (using False Discovery Rate (FDR)-corrected $p$-values < 0.05) to metabolites involved in glycolysis, dihydroxyacetone phosphate (DHAP) metabolism, β-alanine biosynthesis, and polyamine precursor biosynthesis **(**Figure 35C). There were significant decreases in the abundances of the glycolytic intermediates 2-phosphoglycerate and pyruvate and a continuous increase in the fermentation product lactic acid, potentially from the conversion of pyruvate (Figure 34). These observations were unsurprising, as glycolysis is the primary pathway for CFE reactions to create ATP, and it has previously been shown that most glycolytic enzymes are present in *E. coli*-derived lysates[64-66].

Changes in other glycolytic byproducts involved in DHAP metabolism were more unexpected. DHAP is a glycolytic intermediate that has various routes for conversion: (1) isomerization into glyceraldhyde-3-phosphate to enter glycolysis, (2) conversion into the glycerol-associated metabolites dihydroxyacetone or glycerol-3-phosphate, or (3) breakdown into the highly toxic molecule methylglyoxal that can in turn become the less

toxic metabolites 1,2-propanediol or lactic acid[144]. Two metabolites within this pathway, dihydroxyacetone and glycerol-3-phosphate, accumulated significantly during the CFE reaction. These metabolites are substrates for or products of glycerol, but we did not detect any significant changes to glycerol (Figure 36A). The increase in lactic acid could potentially result from conversion of DHAP to lactic acid due to accumulation of methylglyoxal. Methylglyoxal is known to be highly toxic to cells due to its ability to interact with DNA[145], though it was not identified in this data set. If present in the CFE reaction at appreciable levels, methylglyoxal could interact with template DNA and inhibit expression.

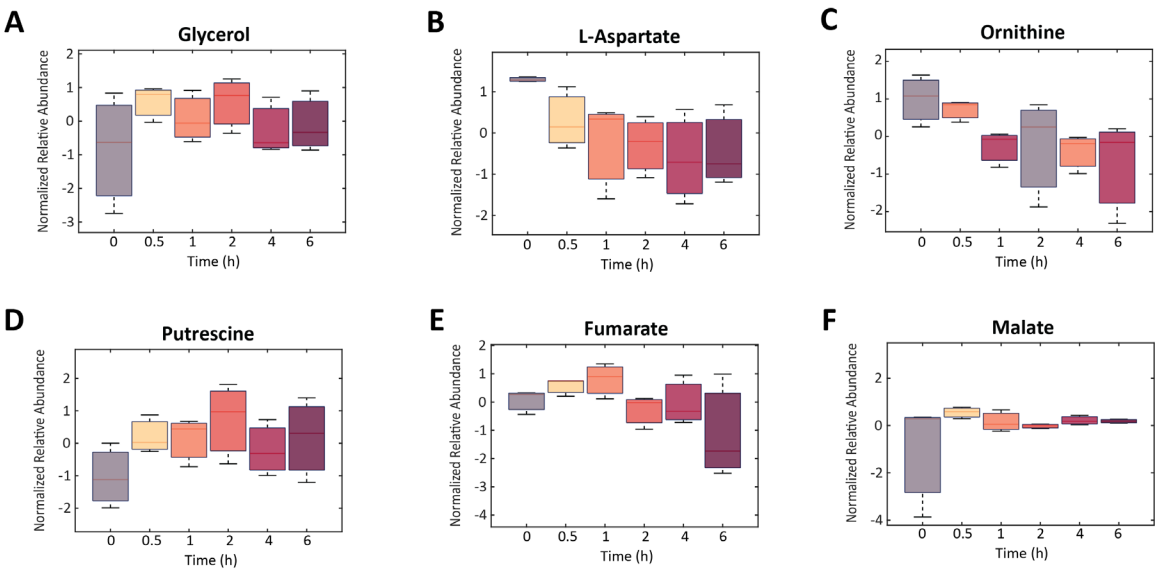


**Figure 36: Relative abundances of individual metabolites in CFE reactions. (A) glycerol, (B) L-aspartate, (C) ornithine, (D) putrescine, (E) fumarate, and (F) malate levels in CFE reaction samples do not significantly change over time. Box and whisker plots depict the normalized peak areas, which are transformed using a generalized logarithm (base 2) and autoscaled. Red lines are the medians, boxes represent the second and third quartiles of values. Error bars represent standard deviation of triplicate reactions.**

Beyond this more central portion of carbon metabolism, some sections of amino acid metabolism also exhibited significant temporal profiles during the CFE reaction,

including β-alanine biosynthesis. β-alanine is the product of L-aspartate, which is synthesized from oxaloacetate (OAA) in the tricarboxylic acid (TCA) cycle. Although L-aspartate did not significantly change (Figure 36B), we found that β-alanine increased at the beginning of the reaction and remained relatively constant after the first hour. This metabolite and its interesting dynamics are notable for a number of reasons. First, it is the precursor to pantothenic acid (vitamin $B_5$) and thus to Coenzyme A, an essential cofactor for many key metabolic pathways including the TCA cycle, fatty acid biosynthesis, and acetyl-CoA production[133, 146]. Second, and perhaps even more noteworthy, we have previously shown that supplementing β-alanine to a CFE reaction increases protein expression[43], indicating that its levels are (either directly or indirectly) important to CFE.

Polyamine biosynthesis was another section of amino acid metabolism with significant temporal profiles. The polyamines putrescine and spermidine are known to be extremely important for processes in living cells due to their key roles in cell-to-cell signaling, cell division, cell motility, and synthesis of DNA and proteins[147]. They are also components of the CFE reaction mixture. Polyamine biosynthesis begins with the molecules L-ornithine or L-arginine, which both can be derived from α-ketoglutarate in the TCA cycle; byproducts of the pathway include urea and methylthio-adenosine. Although ornithine and putrescine remained constant in our measurements (Figure 36C-D) and spermidine was not identified, the precursor metabolites L-arginine and methylthio-adenosine decreased and increased, respectively. Interestingly, the two TCA metabolites we detected (fumarate and malate) did not change significantly (Figure 36E-F), potentially indicating that TCA cycle molecules are not significant precursors for β-alanine or polyamine biosynthesis, likely due to the excess of amino acids in the reaction mixture.

### *4.3.3 The metabolic profile of the lysate alone changes with time*

To begin to pinpoint the source of the metabolic changes that occur in complete CFE reactions, we sought to separately identify metabolic changes in each of the two main constituents of the reaction: the lysate and the reaction mixture. To that end, we measured metabolite profiles in the lysate and in the reaction mixture when separately incubated under otherwise normal reaction conditions. Water was added to the lysate and reaction mixture samples to bring them to the same volume as a CFE reaction, and both were incubated at 37 °C and collected for metabolomics analysis at 0, 1 and 4 h. Data processing yielded 247 and 303 known and unannotated analytes for the lysate and reaction mixture samples, respectively, to be used in further analyses.

Principal component analysis revealed no separation of reaction mixture metabolite profiles at the different timepoints but did yield distinct separation of lysate samples in PC1 **(**Figure 37A-B), consistent with our expectations for both sample types. We expected the small molecules in the reaction mixture to be stable without any enzymes present, and only twelve metabolites (Figure 38) were identified as significantly changing using ANOVA (most of which were not annotated and were likely a result of poor chromatographic peak resolution and different derivatization products). We also expected the enzymes in the lysate to cause changes in metabolite profiles. However, it is worth noting that the lysate had already undergone a "run-off reaction" (to degrade host RNA/DNA) and dialysis[39], making the presence of a lysate metabolome and its potential for significant transformation somewhat surprising.

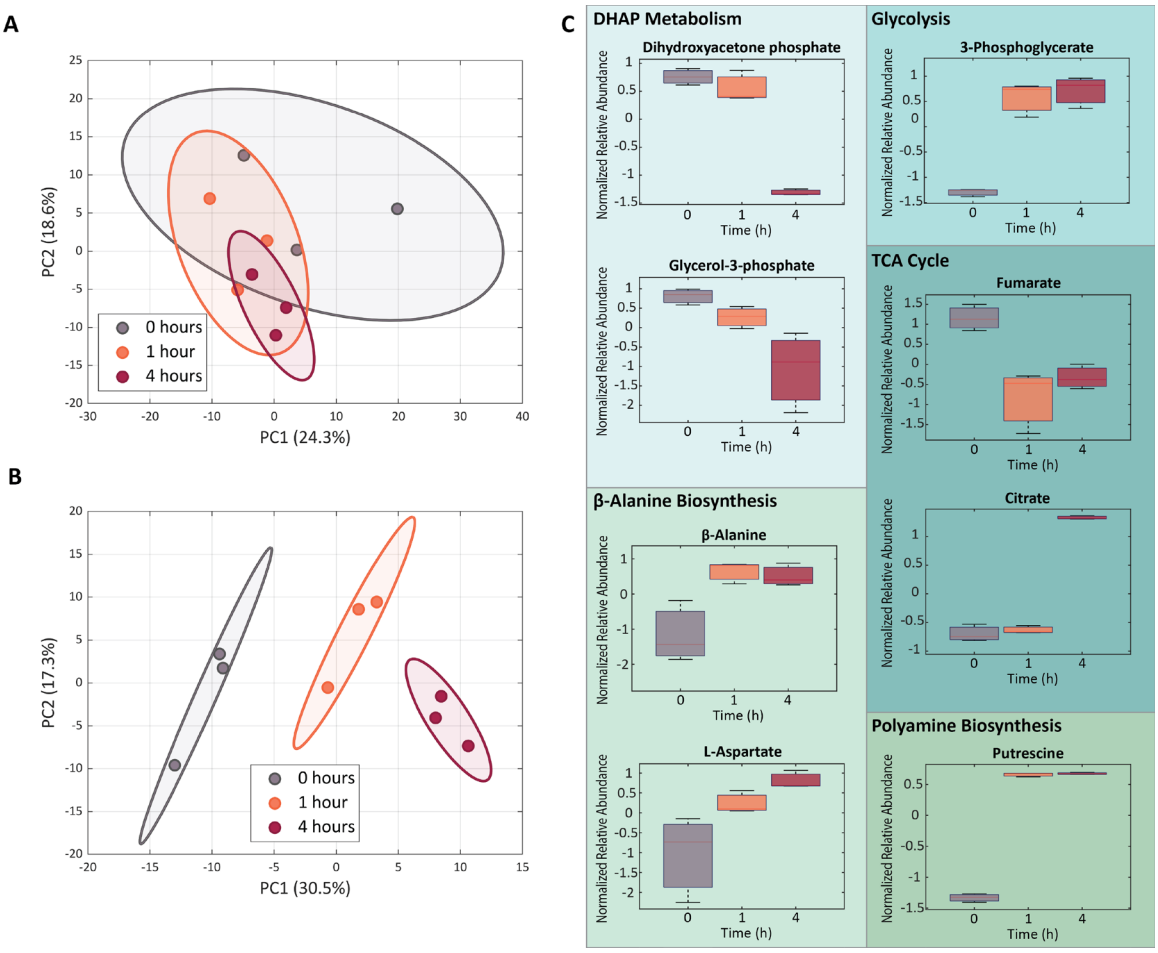


**Figure 37: Metabolic changes in incubated reaction mixtures and lysates.**
**(A) PCA plot for the incubated reaction mixture samples, showing no clustering of or separation between timepoints. (B) PCA plot for the incubated lysate samples, showing distinct separation between metabolite profiles at each time point. For (A) and (B), colored ellipses represented 95% confidence intervals for each group, and the plotted samples are replicate reactions. (C) Metabolites involved in glycolysis, DHAP metabolism, the tricarboxylic acid (TCA) cycle, β-alanine biosynthesis, and polyamine biosynthesis pathways change during lysate incubation. Box and whisker plots depict the normalized peak areas, which are transformed using a generalized logarithm (base 2) and autoscaled. Red lines are the medians, boxes span the second and third quartiles of values. Error bars represent standard deviation of triplicate reactions.**

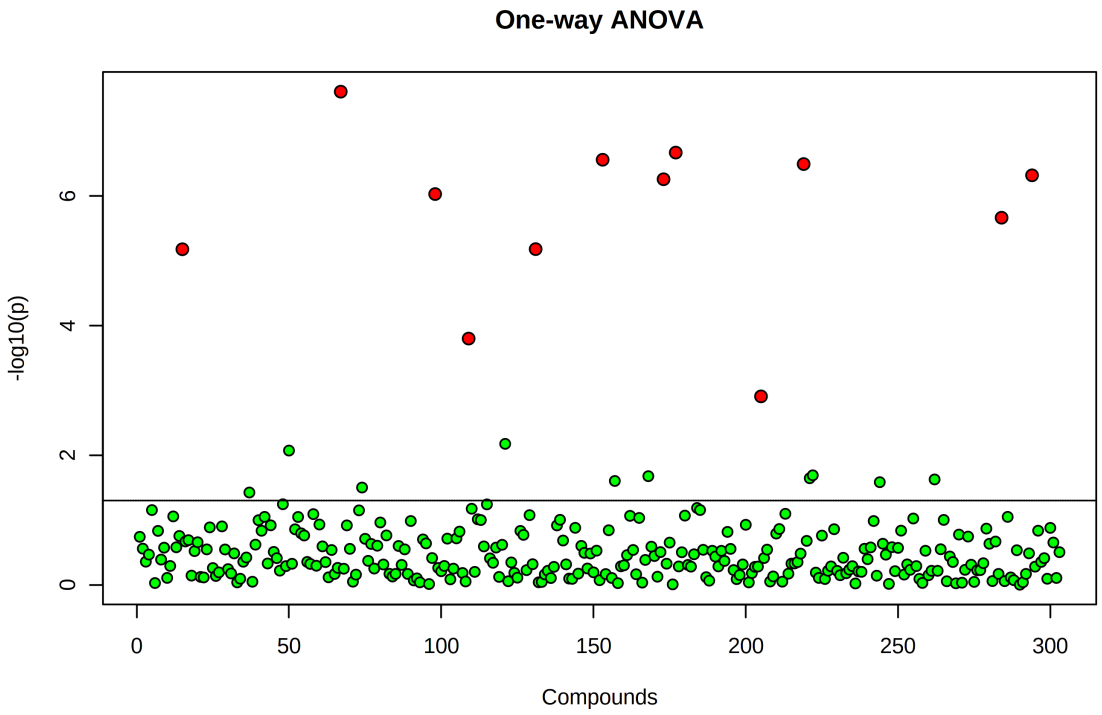


**Figure 38: One-way ANOVA of metabolomics data collected from the incubated reaction mixture samples.**
**Incubation of the reaction mixture in water impacted 12 out of 303 analytes detected. Red circles indicate significantly changing analytes, while green circles represent insignificantly changing analytes. The y-axis indicates raw *p*-values for significance tests for each analyte; the $p = 0.05$ line is drawn. Analyte significance was assessed using a false discovery rate (FDR) corrected *p*-value threshold of 0.05.**

The metabolic pathways with significant changes in the lysate were similar to those in the complete CFE reaction and included glycolysis, DHAP metabolism, β-alanine biosynthesis, and polyamine precursor biosynthesis (Figure 37C). Additionally, molecules in the TCA cycle significantly changed.

Notably, though, many of the metabolites identified as significantly changing in the lysate alone did not have the same temporal trends as in the complete CFE reaction. Glycerol-3-phosphate, citrate, fumarate and putrescine do not trend the same way as in the complete reaction; however, the decreasing trends for glycerol-3-phosphate were not

statistically significant. β-alanine does trend similarly to the complete CFE reaction, but the changes were not statistically significant. Additionally, pyruvate, lactic acid, glycerol, and urea levels did not significantly change over time (Figure 39A-D).

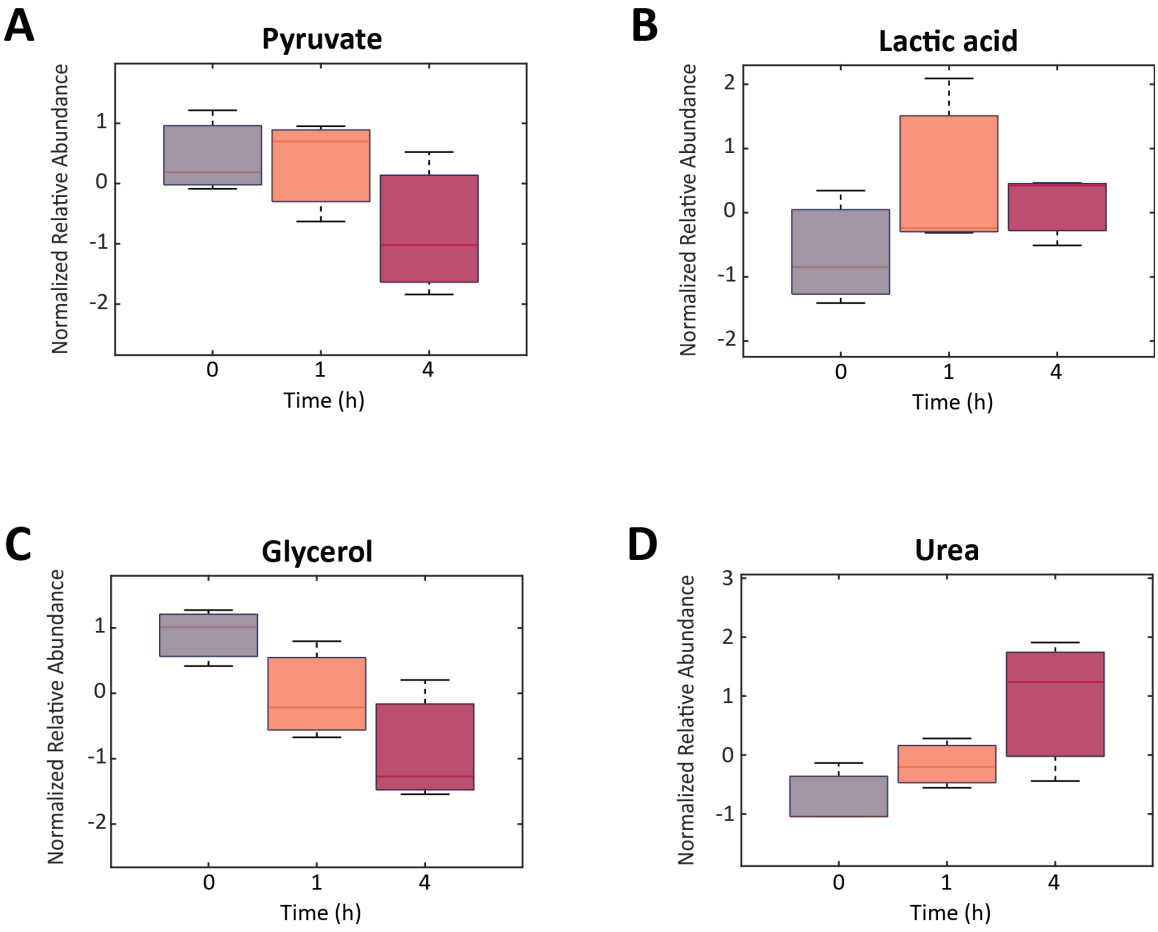


**Figure 39: Relative abundances of individual metabolites in incubated lysate samples. (A) Pyruvate, (B) lactic acid, (C) glycerol, and (D) urea in incubated lysate samples (without reaction mixture or plasmid) do not significantly change over time. Box and whisker plots depict the normalized peak areas, which were transformed using a generalized logarithm (base 2) and autoscaled. Red lines are the medians, boxes are the second and third quartile of values. Error bars represent standard deviation of triplicate reactions.**

Thus, the changes in metabolite profiles during CFE reactions do appear to be attributable to the endogenous metabolic activity of the lysate rather than chemical degradation of the reaction mixture. In fact, similar metabolites are affected in both the

lysate and the complete CFE reaction, suggesting the prominent roles of similar enzymes. However, the different trends in those metabolite profiles between the two cases indicate that the surplus of molecules provided in the reaction mixture fundamentally alters the qualitative impacts of endogenous metabolic activity for those enzymes.

### *4.3.4 Lysate incubation affects protein yield with minor impacts on CFE reaction metabolic state*

We next sought to characterize the relationship between the lysate's initial metabolic state and the productivity and final metabolic state of a CFE reaction. Previously, we demonstrated that pre-incubating the lysate with reaction mixture at 37 °C for 8 h prior to DNA template addition substantially decreased protein output[43], suggesting that the endogenous metabolism of the system may affect its productivity. Since we now know (Figure 35C and Figure 37C) that the metabolic trends of the lysate's endogenous enzymes can be quite different depending on whether the reaction mixture metabolites are present, this led to the question of whether pre-incubation of lysate without the reaction mixture would affect the productivity or final metabolic state of a CFE reaction.

Accordingly, we incubated the lysate without any reaction mixture for six hours at 4, 25, or 37 °C and then used these pre-incubated lysates in a CFE reaction producing GFP from pJL1s70. We selected the incubation time of 6 hours to be consistent with the timescale of protein production (Figure 35), allowing sufficient time for endogenous metabolic activity to approach completion. We found in small-volume microwell plate experiments that lysate pre-incubation at 25 or 37 °C, but not at 4 °C, resulted in a substantial reduction in GFP production (Figure 40). We then used larger-volume reactions

for fluorescence and metabolomics analysis at 0, 1 and 4 hours after the start of the reaction with pre-incubated lysate. After data processing, these metabolomics measurements yielded 424 known and unannotated analytes used in further analyses.

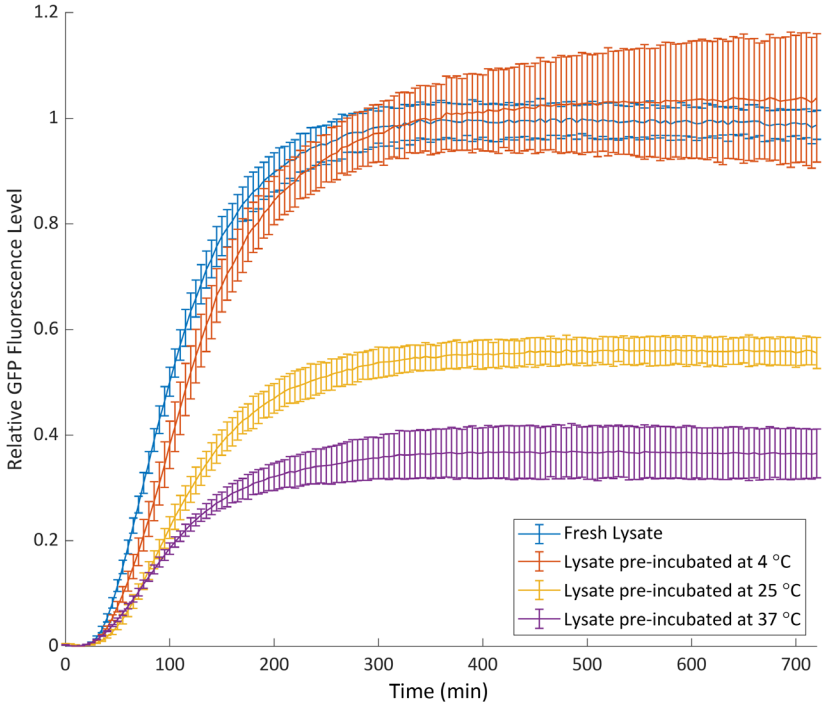


**Figure 40: GFP production from CFE reactions using lysates that were fresh or pre-incubated for 6 h at 4, 25, or 37 °C.**
**Fluorescence was measured every 5 minutes over 10 h of incubation with excitation and emission wavelengths of 485 and 510 nm, respectively, with a gain of 70. Each reaction was 10 µL in volume with 12 nM pJL1s70 plasmid. Error bars represent standard deviation of triplicate reactions.**

As seen in Figure 41A, CFE reactions in the larger-volume format also produce significantly less GFP when lysates are pre-incubated at 25 or 37 °C compared to at 4 °C. Multivariate analysis with PCA indicated differences in metabolic profiles between the three pre-incubation temperatures before the reaction started but a convergence of the overall metabolic profiles by the final timepoint (Figure 41B), a contrast to their divergence in protein productivity. With univariate analyses, we found that most metabolic dynamics

in the reactions with pre-incubated lysates were the same as in reactions with fresh lysate (Figure 41C). For example, temporal trends in glycolysis and TCA cycle metabolites were independent of lysate pre-incubation temperature and identical to the fresh CFE reaction, even though levels of some of these molecules had changed during lysate incubation.

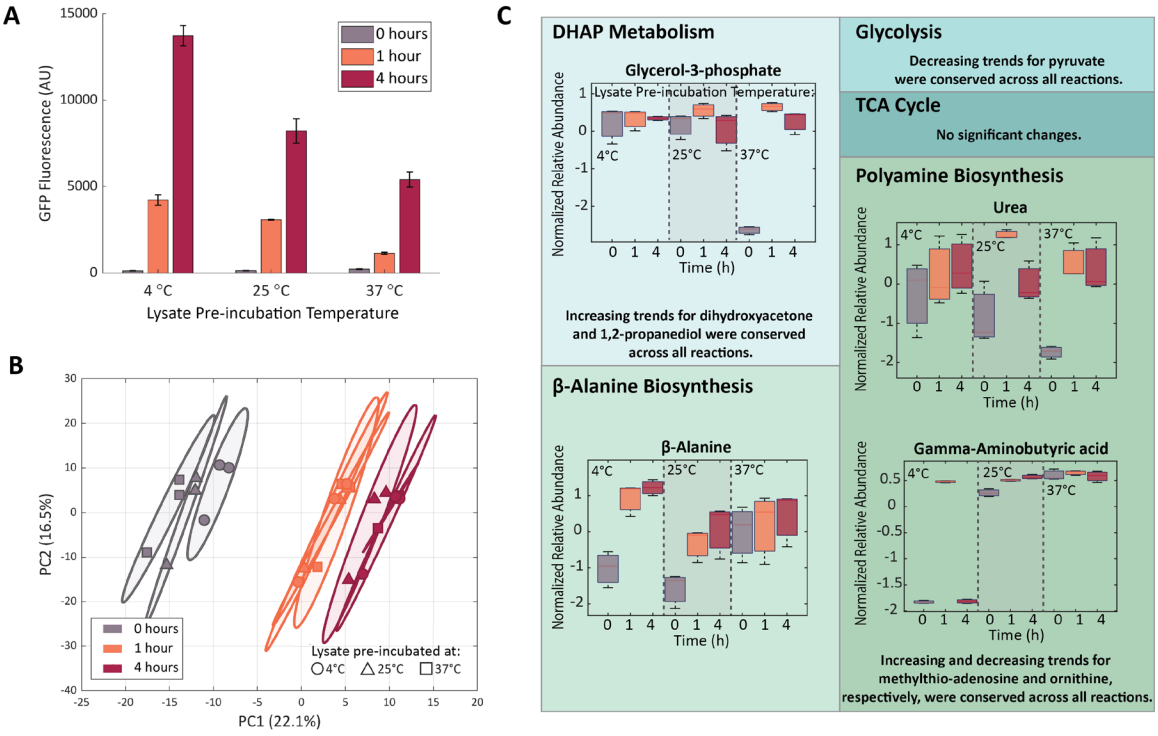


**Figure 41: GFP production and metabolic changes in CFE reactions using lysates pre-incubated for 6 h at 4, 25, or 37 °C.**
**(A) Increasing lysate pre-incubation temperature results in decreased GFP output. Error bars represent standard deviation of triplicate reactions. (B) Principal component analysis of metabolite profiles collected at different reaction timepoints. The samples collected at the start of the reaction separate based on lysate pre-incubation temperature, but the metabolic profiles converge as the reaction progresses. Colored ellipses represented 95% confidence intervals for each group, and the plotted samples are replicate reactions. (C) Only a few metabolites involved in DHAP metabolism, β-alanine biosynthesis, and polyamine biosynthesis pathways had different profiles for different lysate pre-incubation temperatures. Box and whisker plots depict the normalized peak areas, which are transformed using a generalized logarithm (base 2) and autoscaled. Red lines are the medians, boxes span the second and third quartiles of values. Error bars represent standard deviation of**

**triplicate reactions. Metabolites that changed significantly with time but consistently across sample groups are noted with text.**

However, a few molecules in DHAP metabolism, β-alanine biosynthesis, and polyamine precursor biosynthesis had different profiles for different lysate pre-incubation temperatures. Glycerol-3-phosphate levels stayed relatively constant when lysates pre-incubated at 4 and 25 °C were used, whereas with 37 °C pre-incubated lysates this molecule accumulated, indicating it is produced during the CFE reaction despite its downward trend during high-temperature pre-incubation. β-alanine levels increased over time for reactions with 4 and 25 °C pre-incubated lysates (similar to a fresh CFE reaction), but with 37 °C pre-incubated lysates the elevated initial concentration of β-alanine does not appreciably increase, suggesting that β-alanine biosynthesis is largely completed after 37 °C pre-incubation. Urea and gamma-aminobutyric acid (GABA, a putrescine degradation product) were also affected by increasing pre-incubation temperatures, with more accumulation at higher and lower temperatures, respectively. (Putrescine was detected but did not significantly change across samples, potentially due its abundance in the reaction mixture **(**Figure 42).) We also performed ANOVA2 to quantify the impact of lysate pre-incubation temperature on the metabolite profiles. We found that reaction time is the most common significant effect across the measured metabolites, though (consistent with the preceding discussion) there are also significant interaction effects (Figure 43).

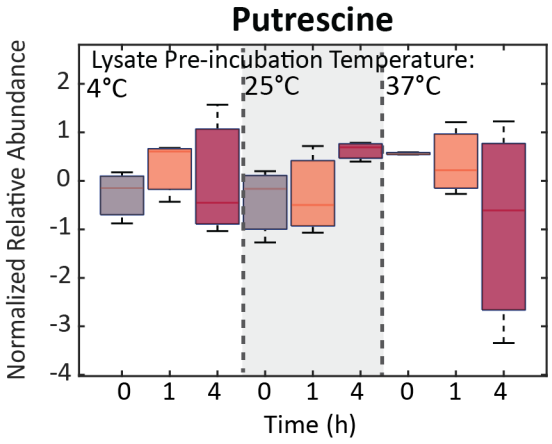


**Figure 42: Relative abundances of putrescine in CFE reaction samples with lysates pre-incubated at different temperatures.**
**Box and whisker plots depict the normalized peak areas, which are transformed using a generalized logarithm (base 2) and autoscaled. Red lines are the medians, boxes span the second and third quartiles of values. Error bars represent standard deviation of triplicate reactions.**

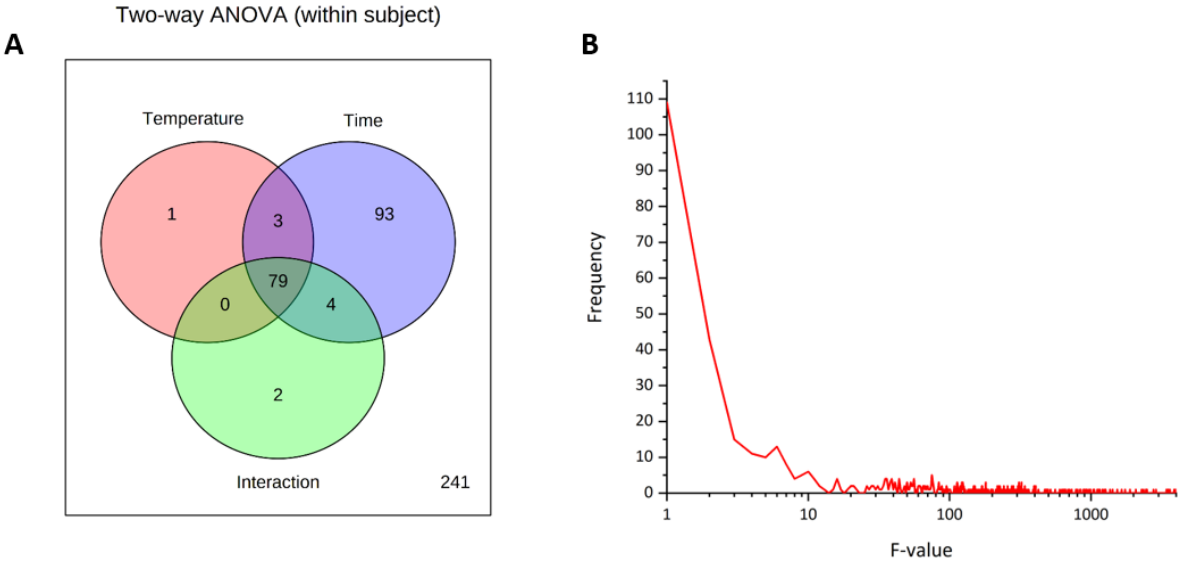


**Figure 43: Two-way ANOVA and the distribution of f-values of metabolomics data collected from CFE reactions run with lysates pre-incubated at different temperature.**
**In (A), numbers represent counts of metabolites that have significant effects for each factor, assessed using FDR corrected *p*-values (<0.05). In (B), bin values of 1 were used to determine f-value frequency. The majority of f-values are between 0 and 9.**

Although pre-incubating lysates at different temperatures significantly affects CFE reaction protein production, the changes in metabolic dynamics caused by the use of these different lysates are small compared to the overall metabolic changes during the course of a reaction. While it was perhaps surprising that the final metabolic states of these systems with such different expression levels seem so similar, the small metabolic changes may very well still be a root cause of the changes in expression, especially since some of these molecules, such as polyamines and β-alanine, are known to have key roles in protein production. This hypothesis is supported by multiple pieces of evidence from the literature. First, the loss of protein expression with pre-incubation seems likely to be either directly or indirectly metabolic in its cause: expression in CFE reactions can be maintained for 16 to 24 hours by use of dialysis reactors[50, 148], suggesting that expression machinery is not likely degrading during the pre-incubation. Furthermore, we previously demonstrated that supplementation of β-alanine, putrescine, and spermidine into CFE reactions can drastically alter protein production, highlighting the impact of metabolite levels on protein output[43]. Nonetheless, proteins that are temperature-sensitive or have short half-lives could be disproportionately affected by pre-incubation, which would not be detected here as enzymatic activity was not directly measured in our experiments. Additionally, the assessment of there only being "small" temperature-related changes in metabolite profiles could be biased by the finite metabolome coverage of GC-MS.

### *4.3.5 Sonication energy input significantly impacts protein production and metabolic profiles*

We next sought to further explore how susceptible endogenous lysate metabolism is to changes in its initial metabolic state via alterations to lysate preparation. Our previous

studies of CFE metabolism assessed the impact of multiple steps in lysate preparation, including growth of the starter culture with or without glucose and dialysis of the lysate[43]. However, we had yet to explore one variable known to have a significant impact on lysate productivity: the sonication energy input to lyse the cells.

While sonication energy input should not impact the initial metabolite profile in the lysate, its impact on expression may be correlated with or mediated by metabolic changes, so we sought to characterize endogenous metabolism in lysates made with different sonication energies. While the exact number varies between operators due to differences in technique, a typical sonication energy input for cell lysis in our group is 250-300 J; lower input energies can be used, but may reduce lysis efficacy due to an increase in intact cells and a lower total *E. coli* protein concentration in the crude lysate[41]. We prepared lysates using sonication energies of 25, 100, and 300 J and used them for assembly of CFE reactions with the reaction mixture and the template DNA pJL1s70. Samples for fluorescence measurements and metabolomics analysis were collected at 0, 1, 4, and 12 h. Processed metabolomics measurements yielded 351 annotated and unannotated analytes for further analysis.

Different sonication energies did in fact lead to different protein yields and metabolite profiles. An energy input of 100 J unexpectedly resulted in a higher protein yield than either 25 J or 300 J at 1 h and 4 h (Figure 44A). While the 25J condition had lower expression than the 300 J condition at 1 h, the 25 J condition surpassed the 300 J condition at 4 h and made comparable amounts of protein to the 100 J condition by 12 h. This temporal difference in expression profiles was reproducibly observed, but given the known operator-specific aspects of sonication energy protocol optimization, we refrain

from interpreting too much from the quantitative details and instead focus on the three conditions as generally indicative of different degrees of lysis and different expression efficiency. The stark differences in protein yield reflected in those measurements were evident in our metabolomics data when analyzed using PCA (Figure 44B), with significant separation of the conditions at almost all timepoints, indicating that different sonication energy inputs lead to fundamentally different endogenous CFE metabolic dynamics. Notably, the 100 and 300 J reaction samples are closer to each other in PCA space than the 25 J reactions at each time point, suggesting greater similarities in their metabolic profiles.

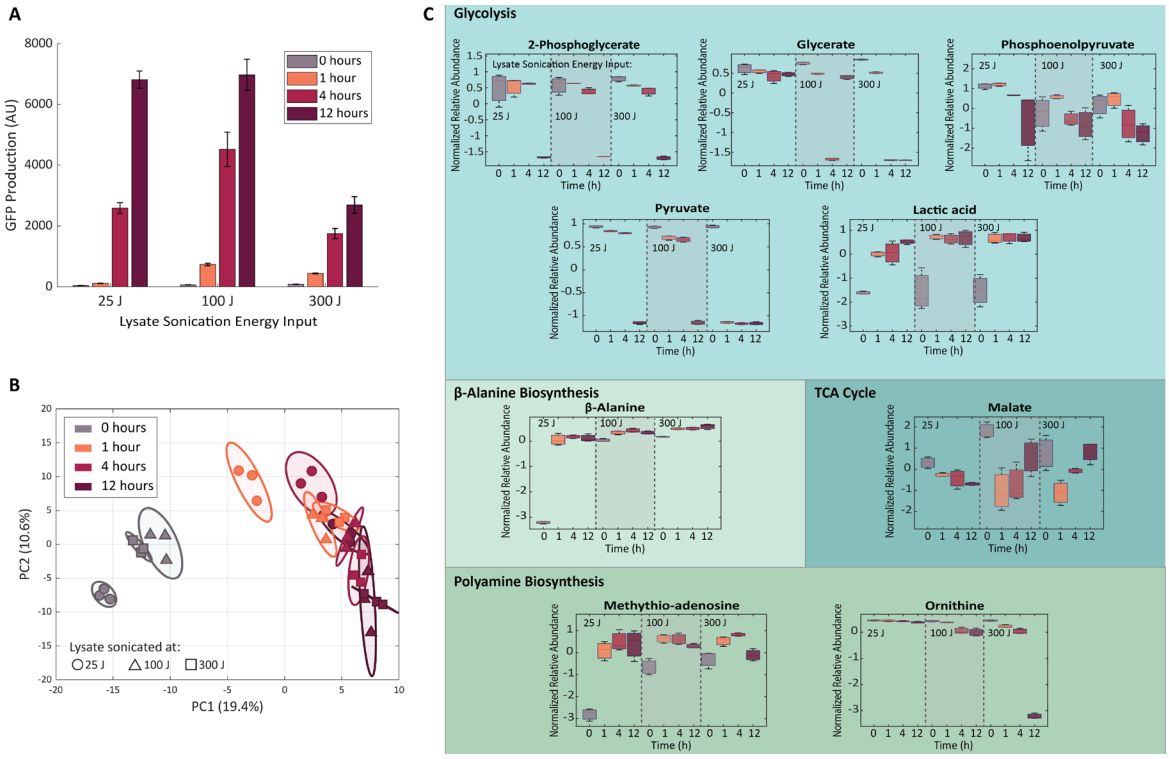


**Figure 44: GFP production and metabolic changes in CFE reactions using lysates sonicated with different energy inputs.**
**(A) Reducing lysate sonication energy input from 300 J to 25 or 100 J significantly improves protein production expression in CFE reactions. Error bars represent standard deviation of triplicate reactions. (B) Principal component analysis of metabolite profiles from reactions using differently sonicated lysates. The different lysate reactions separate in PC1 at 0, 1, and 4 h timepoints. Colored ellipses represented 95% confidence intervals for each group, and the plotted samples are replicate reactions. (C) Metabolites involved in glycolysis, the TCA cycle, β-alanine**

**biosynthesis, and polyamine biosynthesis pathways are prominently affected by sonication energy input. Box and whisker plots depict the normalized peak areas, which are transformed using a generalized logarithm (base 2) and autoscaled. Red lines are the medians, boxes span the second and third quartiles of values. Error bars represent standard deviation of triplicate reactions.**

Univariate (ANOVA) analysis again provided additional context for the multivariate results. 100 and 300 J reactions behaved like our complete CFE reactions and had almost identical metabolic behavior, with few exceptions (Figure 44C). The 25 J reactions appeared to have slower glycolytic activity, as evidenced by the smaller changes between 0 h and 4 h compared to the 100 J and 300 J reactions. Interestingly, the 25 J reactions had lower initial abundances of malate, β-alanine, and methylthio-adenosine, although the trends of all but malate remained the same across sonication energies. Glycerol-3-phosphate, fumarate, succinate, and putrescine were present in the data set and remained constant for all reactions (Figure 45A-D).

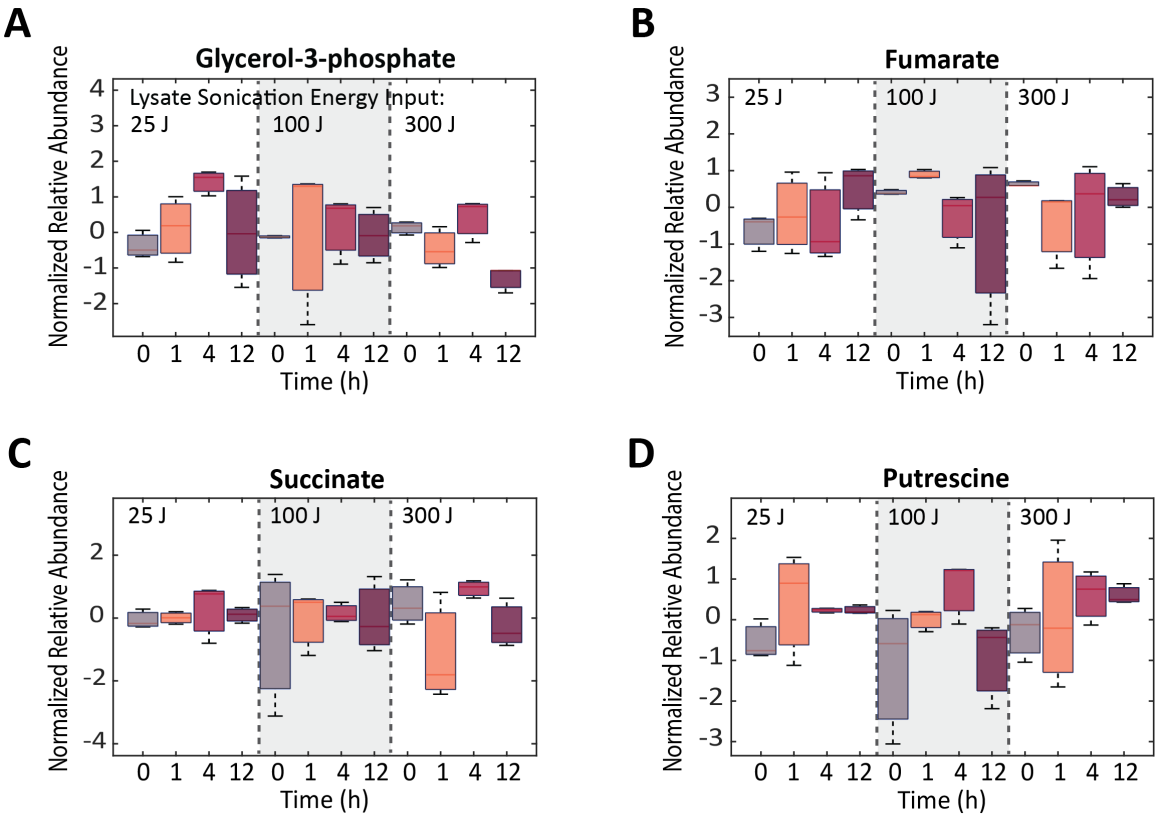


**Figure 45: Relative abundances of individual metabolites in CFE reactions using differently sonicated lysates. (A) Glycerol-3-phosphate, (B) fumarate, (C) succinate, and (D) putrescine do not significantly change over time. Box and whisker plots depict the normalized peak areas, which were transformed using a generalized logarithm (base 2) and autoscaled. Red lines are the medians, boxes are the second and third quartile of values. Error bars represent standard deviation of triplicate reactions.**

ANOVA2 results highlight the importance of time and interaction effects over sonication energy input alone, indicating that (similar to lysate pre-incubation) the initial metabolic differences between lysates with different sonication energies are ultimately blunted over time (Figure 46). While the quantitative metabolic dynamics across these lysates differ, they are often variations on similar trends but to different degrees. This may be a result of the change in total protein content (rather than specific activity) caused by altering sonication energy input. Even though the lysates for the sonication energy input experiment were derived from a different strain and prepared by a different operator than

in the previous figures, their metabolic dynamics were similar to the previous results; this further highlights the resiliency of CFE reaction metabolism to external perturbations, which is a particularly salient feature given the well-known operator dependence of CFE results.

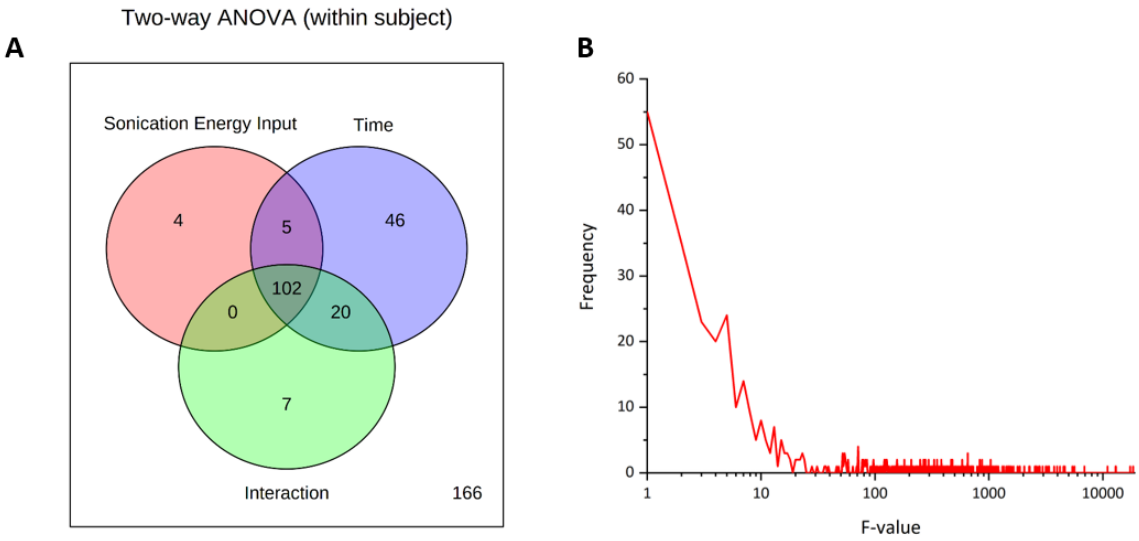


**Figure 46: Two-way ANOVA and the distribution of f-values of metabolomics data collected from CFE reactions using lysates sonicated with different energy inputs. In (A), numbers represent counts of metabolites that have significant effects for each factor, assessed using FDR corrected *p*-values (<0.05). In (B), bin values of 1 were used to determine f-value frequency. The majority of f-values are between 0 and 15.**

We note that because the stability and function of proteins are known to be affected by sonication energy input, proteins involved in transcription and translation could have been affected in our experiment, though they were not directly measured. The metabolic behaviors observed here, along with the unique temporal expression dynamics at different sonication energies, indicate the likely complex interdependencies between proteins and metabolites, and between expression and metabolism, in CFE systems.

### *4.3.6 Targeted enzyme supplementation has minor effects on protein yield and metabolic profiles*

With this additional support for the idea that endogenous metabolism is connected to CFE productivity and potentially robust to changes in the initial lysate's metabolic state, we sought to assess the impact on productivity when some metabolic fluxes are perturbed or rerouted, which could perhaps cause larger-scale metabolic perturbations than single metabolite supplementation. We selected five cytosolic enzymes to supplement in CFE reactions based on our findings to this point and their known importance in central carbon and amino acid metabolism: three glycolytic enzymes (phosphoglycerate kinase (Pgk), pyruvate kinase (PykF), and lactate dehydrogenase (LdhA)) and two TCA cycle enzymes (citrate synthase (GltA) and isocitrate dehydrogenase (Icd)) (Figure 34).

We initially sought to supplement the enzymes via expression from a plasmid; however, we found that including additional plasmid DNA to produce the enzyme at the same time as the GFP reporter in a CFE reaction confounded results (Figure 47). Extra DNA, regardless of sequence and gene products, increased GFP production in these experiments. To avoid this confounding effect, we instead supplemented the reaction with purified enzyme. Before performing metabolomics analysis, we identified the optimal concentration ranges for expression enhancement for each enzyme in small-volume reactions (Figure 48). Only two of the five enzymes (GltA and LdhA) caused improved endpoint GFP production in the CFE reactions when supplemented at their optimal concentrations.

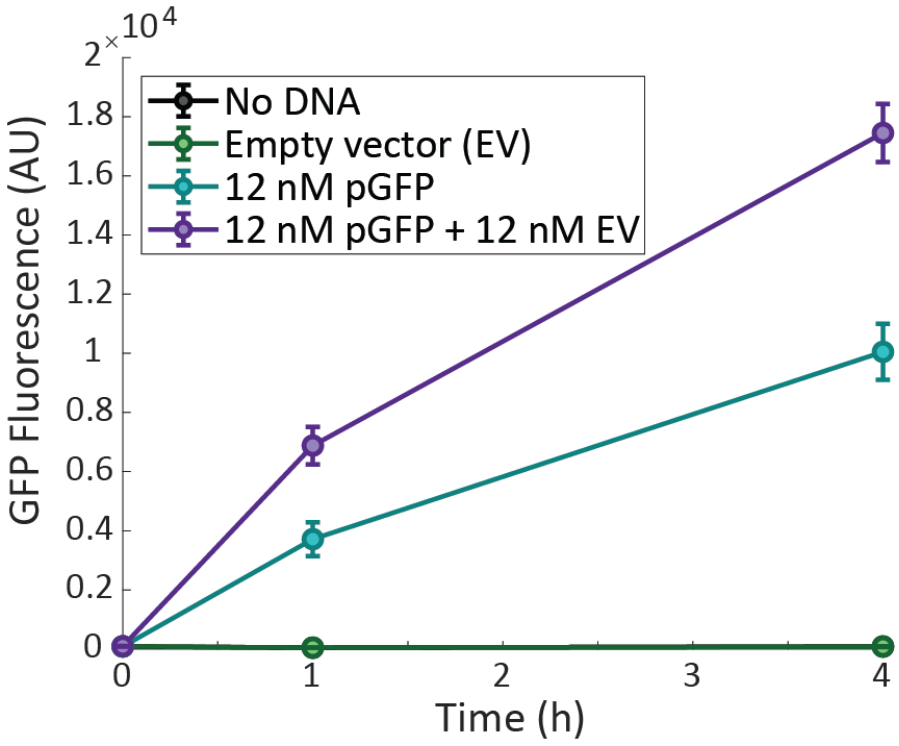


**Figure 47: GFP production changes for reactions run with additional, non-reporter plasmids.**
**pGFP is pJL1s70, with expression controlled by a standard *E. coli* $\sigma^{70}$ promoter. Error bars represent standard deviation of triplicate reactions.**

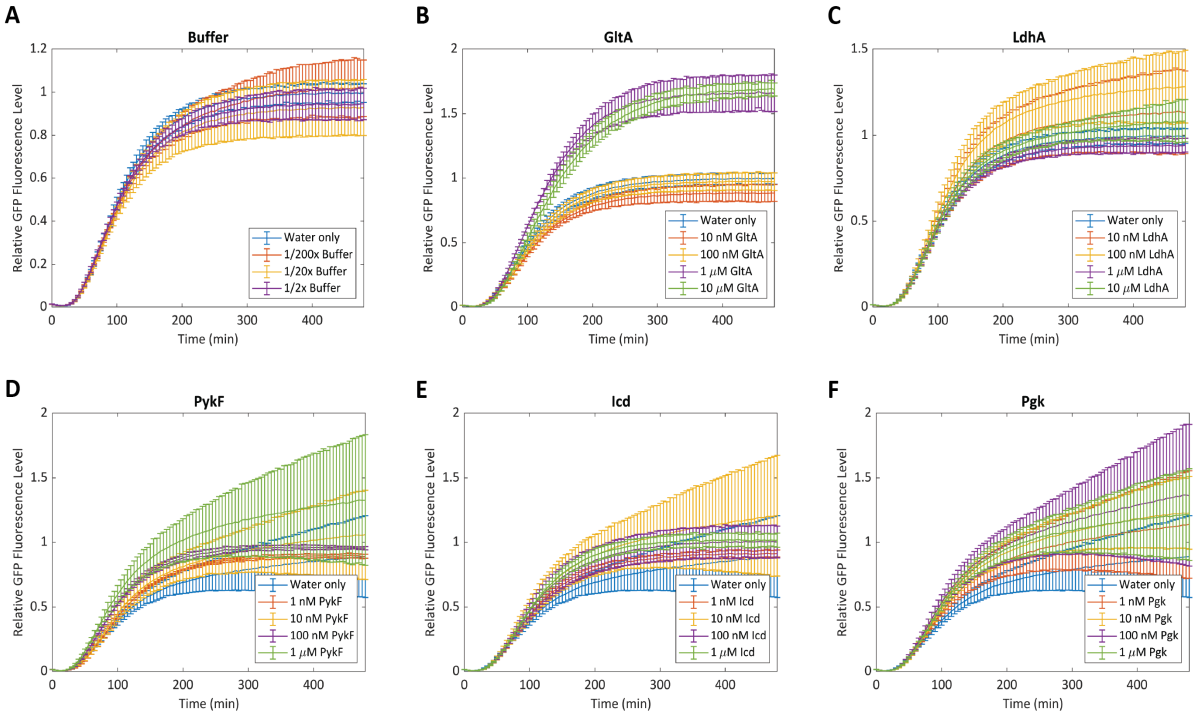


**Figure 48: Optimization of enzyme supplementation levels in a CFE reaction. Different concentrations of (B) GltA, (C) LdhA, (D) PykF, (E) Icd, and (F) Pgk were added to CFE reactions and fluorescence was measured over 8 hours. Protein storage buffer at dilutions equivalent to those that would be added during enzyme supplementation were also tested (A) to control for potential protein expression improvements due to storage buffer components. 1 µM GltA and 100 nM LdhA statistically significantly improve endpoint GFP production. GFP expression was controlled by a standard *E. coli* $\sigma^{70}$ promoter (pJL1s70). Error bars represent standard deviation of triplicate reactions.**

We selected GltA, Pgk, and LdhA for metabolomics analysis of supplemented reactions based on their metabolic pathway diversity (they come from the TCA cycle, glycolysis, and fermentation, respectively) and the preliminary evidence of improved GFP production for GltA and LdhA. Each larger-volume reaction was supplemented with the optimized concentration of the selected enzyme (1 µM GltA, 100 nM LdhA,10 nM Pgk); control reactions were supplemented with enzyme buffer with no enzyme. Samples were collected for fluorescence and metabolomics analysis at 0, 1, and 4 hours after reaction

assembly. After data processing, these metabolomics measurements yielded 331 known and unannotated analytes used in further analyses.

Only GltA supplementation yielded significantly improved GFP expression in the large-volume reactions (Figure 49A), though supplementation with the other enzymes trended in the same direction. (Differences in protein expression between small- and large-volume reactions is consistent with literature reports[149].) Multivariate analysis via PCA yielded no clear separation of samples at each timepoint (though the GltA samples showed a small amount of separation at 0 hours), indicating that supplementation of these enzymes only minimally impacts the metabolic profile despite their known importance in carbon metabolism (Figure 49B).

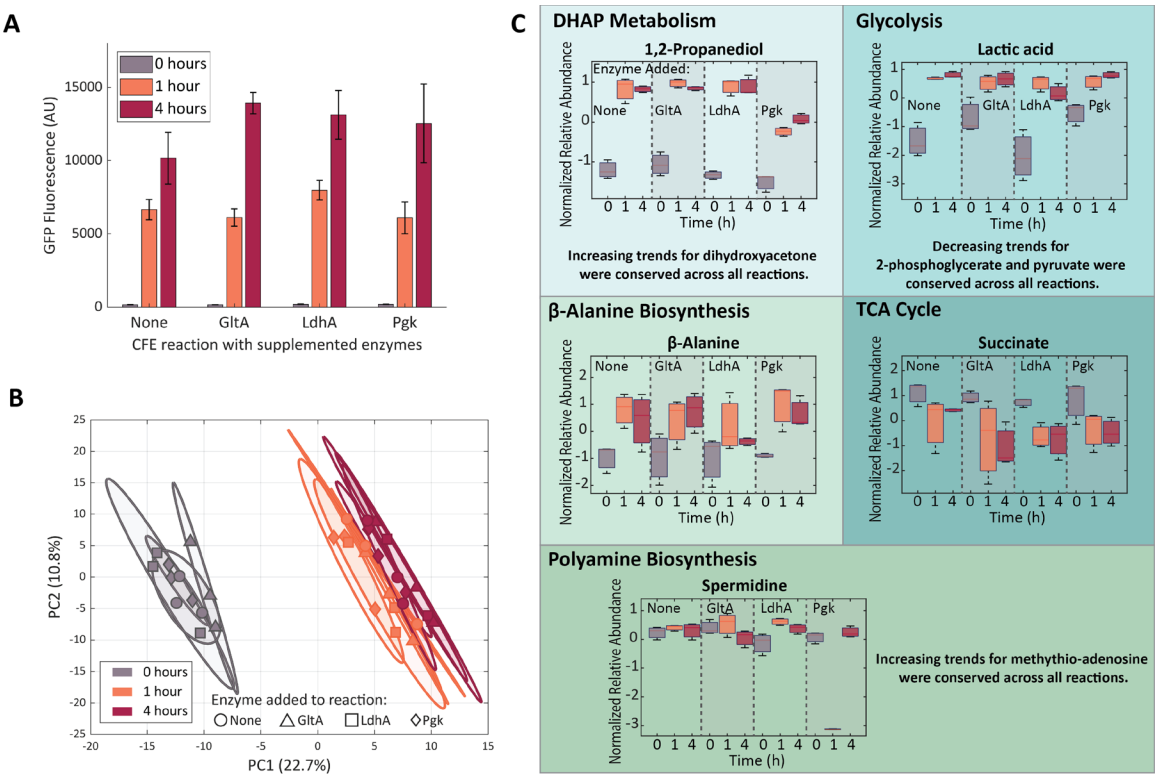


**Figure 49: GFP production and metabolic changes in CFE reactions supplemented with the enzymes GltA, LdhA, or Pgk.**
**(A) Though all three supplemented reactions trended towards increased GFP expression, only GltA supplementation yielded a statistically significant increase at 4 h. Error bars represent standard deviation of triplicate reactions. (B) Principal component analysis of metabolite profiles collected at different reaction timepoints. The supplementation conditions do not clearly separate at any timepoint. Colored ellipses represent 95% confidence intervals for each group, and the plotted samples are replicate reactions. (C) Only a few metabolites involved in glycolysis, DHAP metabolism, the TCA cycle, β-alanine biosynthesis, and polyamine biosynthesis pathways were affected by enzyme supplementation. Box and whisker plots depict the normalized peak areas, which are transformed using a generalized logarithm (base 2) and autoscaled. Red lines are the medians, boxes span the second and third quartiles of values. Error bars represent standard deviation of triplicate reactions. Metabolites that changed significantly with time but consistently across sample groups are noted with text.**

On a univariate level, while most metabolite levels were not affected by enzyme supplementation, there were five metabolites within DHAP metabolism, glycolysis, TCA cycle, β-alanine production and polyamine biosynthesis with notable (though often subtle) changes for different supplemented enzymes (Figure 49C). The DHAP breakdown product

1,2-propanediol increased for all reactions, but to a lesser extent for reactions with Pgk, potentially indicating some rerouting of metabolic flux from DHAP metabolism into glycolysis. Reactions with LdhA had lower levels of lactic acid at 4 hours than the control reaction. Although one may have expected larger changes to lactic acid production for reactions supplemented with LdhA, some of the lactic acid pool may come from conversion of methylglyoxal in DHAP metabolism. Succinate levels also change with enzyme supplementation (though with similar profiles across all three enzymes), and supplementation of LdhA and Pgk both affected biosynthesis of the polyamine spermidine. For GltA supplementation (again, the only one to yield significant increases in GFP expression), only one metabolite (succinate) was notably changed compared to the control. Interestingly, β-alanine levels were largely unaffected, even though one might have anticipated changes based on GltA's consumption of OAA. β-alanine profiles may have remained unchanged due to the excess of amino acids (specifically L-aspartate) supplied in the reaction mixture.

Taken together, the impacts of enzyme supplementation on the metabolic state of the CFE reactions is comparatively small (though again, one should note the caveat that GC-MS is not well-suited to measuring all classes of metabolites, so there may metabolites we did not measure with more clear differentiation between conditions). While analysis with ANOVA2 suggests that there are more statistically significant enzyme supplementation effects than what is evident from visual inspection or one-way analyses, time still has the broadest effect on metabolite profiles **(**Figure 50) and the ANOVA2-significant enzyme effects are still quite subtle. Although one may have expected more substantial changes to metabolic activity due to enzyme supplementation, our results suggest a certain degree of

resilience of CFE metabolism, perhaps a result of the allosteric regulation that allows living *E. coli* cells to maintain metabolite homeostasis[150]. This overall metabolic resilience only further highlights the importance of using metabolism as a guide or target for optimization of CFE systems, as it is a prominent force with a substantial impact on total protein expression.

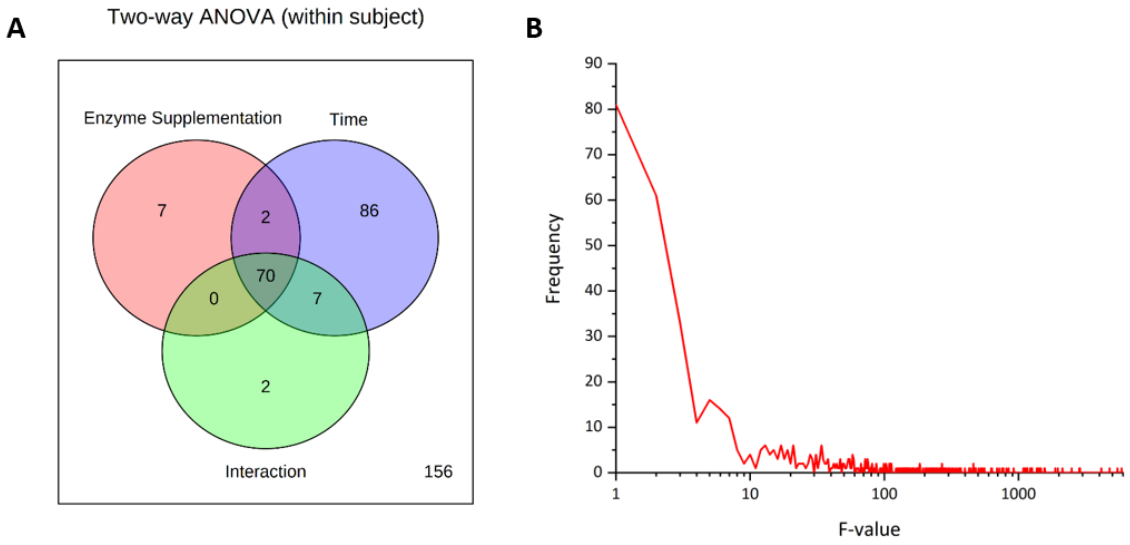


**Figure 50: Two-way ANOVA and the distribution of f-values of metabolomics data collected from CFE reactions run with supplemented enzymes.**
**In (A), numbers represent counts of metabolites that have significant effects for each factor, assessed using FDR corrected *p*-values (<0.05). In (B), bin values of 1 were used to determine f-value frequency. The majority of f-values are between 0 and 7.**

## 4.4 Conclusion

In summary, we identified temporal changes in the small molecules within a CFE reaction via metabolomics using GCxGC-MS analysis, linking them to key areas of metabolism. By performing a deeper characterization of the metabolic dynamics of these systems, we confirmed our findings in the previous chapter that metabolites involved in β-alanine/pantothenate biosynthesis and pyruvate metabolism change during a CFE reaction

as well as discovered that metabolites in other areas of metabolism change including glycolysis, DHAP metabolism, the TCA cycle, and polyamine precursor biosynthesis. We also dissected the contributions of the lysate and reaction mixture to the metabolic changes in a complete CFE reaction, confirming that the lysate, not the reaction mixture, plays a significant role due to its the endogenous metabolic activity. However, the reaction mix alters the initial metabolic state of the reaction and can impact the direction in which metabolites are utilized.

Because we found in the previous chapter that incubation of the reaction without plasmid DNA negatively affects protein yield (Figure 28), we decided to further explore this by performing metabolomics analysis on pre-incubated lysates that were subsequently used in CFE reactions. Although we found that most metabolic changes in the reactions with pre-incubated lysate were the same as those in reactions with fresh lysate, a few molecules involved in DHAP metabolism, β-alanine biosynthesis, and polyamine precursor biosynthesis changed, which is notable because we were able to show in the previous chapter that supplementation of the polyamine putrescine as well as β-alanine can improve reaction performance (Figure 32).

We then decided to determine how susceptible endogenous lysate metabolism is to changes in its initial metabolic state via alterations to lysate preparation method. Because we previously assessed the impact of growth of the starter culture with or without glucose and dialysis of the lysate, we chose to explore how changes to lysate sonication energy input alter protein yield putatively via metabolic changes. Despite major changes to protein output, there were comparatively minor impacts on the overall metabolic profile of the reaction. We then used the information from these studies to selectively target some of the

areas of metabolism (specifically glycolysis, fermentation, and the TCA cycle) via enzyme supplementation. We were able to improve expression, though only in a few cases and it had only minor impact on the metabolic profile of the reaction. Overall, CFE reactions maintain a robust balance of metabolites despite changes to the initial metabolic state of the lysate and enzymatic capacity of the lysate, and subtle and small changes in metabolite levels may play a significant role in determining reaction productivity. Our results highlight the complex, intertwined relationship of metabolism and protein expression as well as their potential as a vehicle for understanding CFE systems to improve their optimization for a broad range of applications, including biosensing.

# CHAPTER 5. CONCLUSIONS AND FUTURE DIRECTIONS

In this thesis, I have used metabolomics to explore key obstacles of bacteria-based whole-cell and cell-free biosensor design. I showed that a better understanding of metabolism can lead to improvements in biosensor engineering. I have identified specific strategies to improve biosensor performance in whole-cell and cell-free systems. However, there still remain important next steps to evaluate the utility of these findings for both whole-cell and cell-free biosensors. In this chapter, I will discuss the novelty and contribution of my thesis research as well as the relevant next steps to expand upon the advances and findings in my work. I will also present more forward-looking directions to further expand the applicability of whole-cell and cell-free bacterial-based biosensors through the use of metabolomics.

## 5.1 Novelty of thesis research

While metabolomics analyses have been widely applied to study *E. coli* cells that have been designed to maximize biochemical production, very few studies have focused on the metabolic impacts of engineering efforts related to biosensor development. In my work, I sought to bridge this gap by exploring the metabolic effects linked to expression and optimization of a well-characterized biosensor reporter system. I characterized the metabolic impacts of a previously-reported strategy to improve carotenoid production, and I was the first to link homocysteine and homoserine accumulation to mevalonate precursor production. Additionally, I developed an easily implementable and effective strategy to mitigate some of the toxicity associated with homocysteine and homoserine and uncovered other metabolites that may also be linked to the toxicity that could be investigated to

potentially reduce cellular stress even further. Overall, this work highlights the value of applying metabolomics to inform strain design for not only lycopene production but in general for whole-cell biosensors.

Despite the major advances in applying CFE systems for biosensor development, there had been no studies before my work that characterized the metabolic profiles of *E. coli*-based CFE systems. Previous research had focused only on the protein synthesis impacts of changes to CFE systems, including alterations to lysate preparation protocols. By directly measuring the metabolite levels within CFE systems, I identified that pantothenate precursors and polyamines are impacted based on lysate preparation method and demonstrated that molecules such as putrescine, homocysteine, and β-alanine can be supplemented into the lysates to improve reaction performance and enable consistent activity across different batches of lysates. The results from this thesis can be directly applied to better engineer biosensors and have already been useful to inform the development of a homocysteine biosensor in our lab, as they have provided novel insights on the endogenous production of homocysteine in CFE reactions. Due to the overall popularity of the lysate methods studied, the findings and proposed supplementation strategies in this work are highly generalizable to a variety of bacterial cell-free systems and are especially useful for biosensor design to help optimize and standardize sensor response.

To enable more informed engineering efforts of CFE systems, I characterized the dynamic small molecule changes in CFE lysates and reactions. I identified active areas of metabolism including glycolysis, DHAP metabolism, the TCA cycle, β-alanine synthesis, and polyamine precursor biosynthesis that consistently change during a CFE reaction,

independent of GFP production, lysate pre-incubation, changes to lysate sonication energy input, and even enzyme supplementation. I demonstrated that these trends are conserved in CFE reactions using lysates derived from different strains of *E. coli* that were prepared by different operators. Because central carbon metabolism should be well-conserved across different strains, these results are broadly applicable to a variety of CFE systems derived from different strains of *E. coli*, and the metabolic trends will likely be conserved across batches prepared by different users. This thorough characterization of the endogenous metabolic processes in CFE systems and the metabolic effects of system perturbations can help inform which metabolic pathways should be targeted for CFE reaction optimization. Overall, this information provided in this thesis can enable more rational and informed engineering strategies for improved bacterial biosensor design as well as CFE system performance in general.

## 5.2 Next steps and future directions for whole-cell biosensors

### *5.2.1 Next steps – Expanding strategies to improve lycopene production and mitigate toxicity*

As shown in this work, the lycopene reporter alone did not substantially impact cellular metabolism, but it also did not produce sufficient levels of lycopene in a short enough timeframe for biosensing applications without further optimization via precursor pathway introduction. Although the precursor pathway could increase the extent and rate of lycopene production, their impacts were highly dependent on when the pathway was induced. Because accumulation of homocysteine and homoserine was found to be linked to the toxicity associated with expression of the precursor pathway at certain times, the

next step for this study should be to further explore strategies to minimize the toxicity associated with these small molecules to potentially allow for optimized carotenoid biosynthesis. Additionally, the molecules ethanolamine and cysteine sulfinic acid accumulated when lycopene was produced for the strains expressing the precursor pathway during overnight growth and should be further investigated to determine their role in cellular toxicity.

Although I have already shown in this work that methionine supplementation can alleviate some of the stress caused by homocysteine and homoserine accumulation, lycopene was not quantified for the strains grown with methionine. A clear next step would be to measure the lycopene production in these strains to determine if methionine supplementation is a valuable approach to improve lycopene synthesis and increase its rate of production, which is necessary to decrease the time needed to generate a visual output. Additionally, methionine supplementation at the start of lycopene production was not sufficient to completely restore normal growth, indicating the potential for further improvement. Supplementation of methionine to cells at the time of induction of the mevalonate pathway may be a better approach to relieve stress and improve lycopene production. Also, isoleucine and leucine have been shown to improve growth of *E. coli* exposed to high levels of extracellular homocysteine[140], so these small molecules should be explored to determine if they are also valuable supplements to improve growth and reporter production.

Because small molecule supplementation efforts may not be sufficient to reduce intracellular stress, genetic perturbations can be explored to determine if they better reroute metabolic flux and reduce homocysteine and homoserine accumulation. To evaluate this,

key enzymes used to metabolize homocysteine and homoserine in precursor producing cells can be overexpressed from plasmids. For homocysteine, the MetE and MetH enzymes could be overexpressed because both are involved in catalyzing the reaction of homocysteine into methionine[151]. For homoserine, the MetA and ThrB enzymes can be used because they catalyze the reactions of homoserine to methionine and homoserine to homoserine phosphate (the threonine biosynthetic precursor), respectively[152].

Deletions of the upstream genes that result in the production of homocysteine (*malY* and *metC*) and homoserine (*metL* and *thrA*) could also be targeted to prevent accumulation. However, this approach is less suitable because these genes have key roles in branched chain amino acid synthesis, and their deletion may have undesirable downstream consequences on amino acid metabolism, introducing more stress to the cells. Clustered regularly interspersed short palindromic repeats (CRISPR) interference (CRISPRi) is an alternative approach that can be used to repress the expression the genes involved in the production of homoserine and homocysteine[153]. CRISPRi requires a catalytically dead variant of the CRISPR-associated (Cas) protein (dCas) and non-coding, custom-designed guide RNAs (gRNAs) that guide dCas proteins to the target gene sequence, and the resulting dCas-gRNA complex sterically hinders transcription of the gene. Gene repression could be induced at specific times when the toxic molecules are known to accumulate, such as during lycopene production, to alleviate cellular stress. Because CRISPRi-based gene repression is tunable and reversible, any stress associated with gene repression can be reduced by restricting available dCas proteins and gRNAs to limit the overall repression.

To validate our findings regarding homocysteine and homoserine and further explore the changes in ethanolamine and cysteine sulfinic acid accumulation, metabolic

flux analysis could be performed to profile the pathways that utilize these molecules, such as the cysteine, methionine, and serine biosynthesis pathways. This type of study tracks and quantifies stable isotope labeled metabolite intermediates to measure metabolic flux[154] and can be used to determine how the fluxes into the amino acid pathways are affected based on induction time of the mevalonate pathway coupled and uncoupled to lycopene synthesis. Overall, this type of experiment can offer more clarity into how the metabolite intermediates in the identified pathways are affected and potentially provide mechanistic insight which could lead to better-informed interventions. This approach could also identify interesting changes in molecules that are known to be toxic, such as buildup of the toxic molecule acetaldehyde from ethanolamine metabolism[155].

### *5.2.2 Future directions – Extending the utility of bacterial biosensors*

Since optimization of the reporter system can be challenging and lead to unforeseen metabolic consequences, it is vital that we improve our understanding on the metabolic impacts of engineering and optimizing reporter systems to engineer these systems more effectively for easier and more rapid biosensor development. This is especially important since researchers often select reporter systems based on their familiarity and experience with the system as opposed to selection based on a systematic investigation of the optimal reporter[11]. Future work for this project should involve expanding our knowledge on the metabolic impacts of other common reporter systems in whole cells and evaluating how typical ways to optimize these systems further affect metabolic state. Specifically, fluorescent, bioluminescent, and colorimetric reporter systems such as GFP, luciferase, and lacZ could be studied. Optimization efforts such as manipulation of plasmid copy number,

promoter strength, and ribosomal binding site strength can also be explored to help inform the selection of the reporter system for specific biosensing applications.

Additionally, our lab has successfully used multiple colorimetric reporter systems to provide visual, semi-quantitative information on the concentration of target analytes for a field-friendly, point-of-care diagnostic test[23]. For this work, three pigment reporters were used (violacein, lycopene, and β-carotene); however, engineering cells to display this behavior was extremely challenging due to metabolite toxicity[53]. As demonstrated in this thesis work, metabolomics is a useful technique to identify toxic metabolites and could be applied to explore the metabolic effects of expressing multiple colorimetric reporter systems. These studies could explore how efforts to tune the response of each reporter system to specific analyte concentration ranges impact metabolism. New strategies to optimize these systems and expand the number of colorimetric reporters that can be expressed in a cell could potentially be achieved to create more informative, field-friendly biosensors with reduced metabolic demand.

Bacterial biosensors also have huge promise in the area of targeted drug delivery, since they can serve as "theranostics" to sense and respond to a particular environment or condition and deliver a therapeutic payload[156]. For example, bacterial biosensors have already been used to detect pathogenic bacteria and respond by producing a pathogen-specific toxin[157] and have even been designed to deliver a cancer-targeting nanobody in response to bacterial cell density near tumors[158]. However, production of these therapeutic molecules can have major impacts on cellular growth and metabolism, limiting the utility of bacteria as theranostics. Metabolomics is an ideal technique to characterize the metabolic effects of therapeutic molecule production in biosensing cells and can even be

extended to evaluate metabolic changes associated with biosensor-host interactions. This approach could reveal underlying metabolic changes that result from therapeutic production and host environment that could be targeted to improve sensing and drug delivery, ultimately enabling the development of bacterial theranostics to target and treat a broad range of diseases.

## 5.3 Next steps and future directions for cell-free biosensors

### *5.3.1 Next steps – Further exploring the importance of identified metabolites and pathways on CFE reaction activity*

For CFE systems, I have identified metabolic pathways involved in glycolysis, DHAP metabolism, the TCA cycle, β-alanine synthesis, and polyamine synthesis as well as homocysteine to change over the course of a CFE reaction and within an incubated lysate. These metabolic pathways and molecules are ideal areas to further explore for efforts to optimize protein production and improve biosensor response.

A relatively straightforward way to do this is to extend the small molecule and enzyme supplementation efforts to encompass different types of metabolites and enzymes. This could be especially effective since both strategies resulted in improved protein production in CFE reactions. For example, because glycolysis and the TCA cycle are important energy-providing pathways for CFE reactions and supplementation of the TCA cycle enzyme GltA improved reaction performance (Figure 49A), these areas could be further targeted to explore their impact on protein production. One class of molecules that could be supplemented to more broadly impact glycolysis and the TCA cycle are known glycolytic and TCA cycle inhibitor metabolites such as citrate[159] and succinyl-CoA[160] and

known activators such as ADP[160] and NADH[161]. Additionally, because the importance of DHAP metabolism intermediates was largely unexplored, supplementation of molecules in this pathway such as DHAP, DHA, and methylglyoxal could help us better understand the effects of these molecules on the activity of CFE reactions.

Although I have already explored how supplementing purified enzymes within glycolysis and the TCA cycle affect CFE reactions, enzymes more specific to β-alanine, polyamine, DHAP, and homocysteine production should also be tested. For example, the enzymes involved in β-alanine (PanD) and putrescine (SpeB and SpeC) synthesis could be added to explore how increasing the availability of these small molecules to CFE reactions affects protein yield. Although there exist numerous enzymes that could be tested, these suggestions can provide a starting point for identifying which enzymes are important and whether enzyme supplementation is more effective at improving protein yield than small molecule supplementation.

Additionally, supplementing small molecules and enzymes at various times throughout the reaction could result in increased productivity and is another area that could be examined. Based on the metabolite dynamics I measured over the course of a CFE reaction, certain molecules or enzymes could be added at the timepoints when specific metabolites are known to be consumed or accumulating to improve protein output. Once the metabolites and enzymes that give the highest protein yields are identified, combinations of these molecules can be pursued to find an optimal mixture.

Another way to target these identified pathways for optimization of CFE systems would be to engineer the starter culture strain to underproduce toxic molecules or to

overproduce beneficial ones. Although this approach would be more challenging than direct supplementation, it could be an effective strategy to reroute the flux within these pathways and could even result in changes to the protein expression levels in the cells, ultimately altering the protein profile of the final lysate. For example, because this thesis work highlighted the importance of putrescine in CFE reactions, putrescine degradation genes such as *speE* and *puuA* genes could be deleted. Genes involved in the other identified areas of metabolism could also be targeted, though extra research should be done in selecting genes to ensure that they are not essential, and their deletion does not cause significant stress to the cells. Metabolic enzymes in these pathways could also be overexpressed in the starter culture strain via plasmid-based gene expression. Like direct enzyme supplementation, this should increase the amount of enzyme in the lysate, but it will also alter metabolite levels and potentially even the broader proteomic profile in the lysate because these enzymes will be expressed and functional during pre-lysis cell growth.

Because the metabolic activity in CFE reactions is controlled by the proteins and enzymes within the lysates, understanding how specific proteins change over time could provide more context for the changes we observe in metabolism. Although some studies have identified some of the proteins native to the lysate[64-66], it is unclear if these enzymes are present throughout the entirety of a reaction or if they degrade over time. To address this question, a multi-omics approach can be taken that combines metabolomics and proteomics to understand how the small molecules and proteins in CFE reactions change over time. This type of study would allow us to uncover links between these different levels of cellular activity, expanding the potential for more targeted reaction optimization efforts.

### *5.3.2 Future directions – Extending the utility of cell-free biosensors*

Because there has been almost no characterization of the metabolic processes in CFE systems up to this point, extending this work to further profile the metabolic behavior of CFE reactions in response to changes to the lysate preparation procedure can uncover new opportunities to optimize protein production. Metabolomics can be extremely useful to explore how changes to starting culture conditions, lysis method, and post-processing affect metabolism. To evaluate simple changes to the starter culture growth method, variables such as temperature, carbon source, and time of collection can be evaluated. Changes to these factors have been shown to affect protein content in *E. coli* cells[162-164], so it is likely that these changes will result in alterations to metabolite and protein content of the lysate. Different types of lysis methods besides sonication can be evaluated and compared such as bead-beating and enzymatic lysis to explore how the different techniques alter the metabolite content in the lysate. To further explore the effects of post-processing, the importance of each individual post-processing step such as the wash, run-off reaction, and dialysis can be explored to understand how each affects CFE reaction productivity and metabolic activity. Ultimately, understanding the impact of alterations to the lysate preparation steps is crucial to address the issues of reproducibility, scalability, and standardization in CFE systems and expand the generalizability of these systems to a wide range of biosensing applications.

Two key factors that can further expand cell-free biosensor applicability are rapid response times and field-friendliness. Although our cell-free systems typically begin to produce protein within the first hour and max out production a few hours later, shortening the time needed to generate a measurable response could be more desirable, especially when designing point-of-care diagnostic biosensors. An important next step to address this

would be to perform metabolomics analysis on the earlier timepoints of the reaction. 10-minute interval samples could be taken over the first hour to explore the metabolic dynamics. This could be an especially interesting area to study since we saw large separation of CFE reaction samples collected at the reaction start and after 30 minutes in PCA space, which suggests that major metabolic changes occurred very early on (Figure 35). Additionally, because we primarily characterized the metabolic dynamics in reactions using lysate ND, it is unknown whether the other types of lysates in CFE reactions have the same metabolic behavior at those time points. Evaluating more and earlier time points can help identify which lysates are more suitable than others for specific performance goals and could enable lysate-specific optimization efforts to decrease sensor response time.

For all of the metabolomics experiments described in Chapters 3 and 4, CFE reactions were run in 1.5 mL microcentrifuge tubes; however, tubes are not necessarily the ideal platform for a diagnostic test. Of particular note, our lab has observed changes in CFE performance depending on the vessel geometry (data not shown), which is thought to be attributable to changes in oxygenation of the liquid reaction volume. Therefore, it may be important to identify if changes to reaction vessel geometry affect CFE reaction metabolism. 200 µL CFE reactions (optimal reaction volume for GC-MS analysis) could be performed in 1.5 ml microcentrifuge tubes, 50 mL centrifuge tubes, and 9-well plates and collected for metabolomics analysis to evaluate how changes in vessel geometry and gas interfacial area alter metabolic behavior. Smaller tubes such as PCR tubes could also be used to explore smaller geometries; however multiple reactions would need to be pooled to have a sufficient sample volume for GC-MS analysis. Such a study could lead to insights on which biosensor vessel geometry yields highest or fastest reaction productivity.

The results from this thesis as well as the metabolomics studies suggested here can provide insights on the metabolites that majorly affect CFE system activity and uncover how they change over the course of a reaction. This could inform the development of novel reaction platforms that continuously and selectively remove metabolites that negatively impact activity as well as supplement beneficial metabolites at appropriate times. For biosensing applications, this could potentially be achieved by using microfluidic devices with channels that selectively draw undesirable metabolites away from the main reaction chamber as they are produced and introduce beneficial molecules into the reaction as they are consumed. Designs such as these could greatly improve the response time of biosensors as well as result in optimized, reproducible reaction performance, which is essential for biosensor commercialization. Larger scale formats could be designed similarly to the continuous CFE systems discussed in Chapter 1, but unlike the continuous CFE systems, this platform would add and remove specific small molecules from the reaction to increase yield while decreasing reaction variability. These redesigned reaction formats could majorly benefit biomanufacturing efforts and could facilitate the production of valuable molecules such as pharmaceutical precursors at industrially-relevant titers while reducing batch to batch variability. Additionally, these efforts could extend the life of CFE reactions and further improve target molecule yield, expanding CFE system's utility in both biosensing and biomanufacturing.

## 5.4 Closing Remarks

In this thesis, I have described my efforts to apply metabolomics to address challenges in *E. coli*-based whole-cell and cell-free systems for biosensor development. In whole cells, I evaluated how expression of a well-characterized colorimetric reporter

system and how optimization of reporter production by introducing precursor pathways can result in undesirable changes to metabolism and negatively impact characteristics of the *E. coli* cells that are important to biosensor applications. Moreover, I identified the molecules that partly attributed to cellular stress and identified a successful strategy to alleviate some of the stress via supplementation.

In CFE systems, I characterized the small molecules in differently prepared cell-free protein lysates and investigated how these different lysates affect protein production and the metabolic profiles of CFE reactions. I also profiled the metabolic dynamics in lysates and CFE reactions and demonstrated how alterations to the starting metabolite compositions of the lysates and reactions affects metabolic behavior. I have identified metabolites involved in various areas of metabolism that change over the course of a CFE reaction and are linked to protein production.

Because both whole-cell and cell-free platforms are widely used for biosensor development, solving the challenges that limit the design of each system is crucial to expand the applicability of bacterial-based biosensors. In this work, I have used metabolomics to explore some of the unique issues in biosensor design for each platform and provide insights into areas of metabolism that can be targeted in efforts to optimize bacterial biosensor performance. This thesis highlights the versatility of metabolomics to uncover the metabolic impacts of biosensor engineering in two unique platforms and lays the groundwork for metabolomics-driven optimization efforts to facilitate biosensor engineering and ultimately broaden the reach of whole-cell and cell-free bacterial-based biosensors.

# APPENDIX A

## A1. GCxGC-MS Methods

### *A.1.1 Auto-sampler Method*

An Agilent 7683 auto sampler was used. Prior to sample injection, three pre-washes were performed with pyridine. Samples were pumped 4 times for thorough mixing and injected using a syringe size of 10 µL with 1 µL of injection volume. Three post-washes of the needle were performed after injection using pyridine.

### *A1.2 GCxGC Method*

An Agilent 7890 gas chromatograph adapted to GCxGC analysis was used. Helium was used as the carrier gas with a corrected constant flow rate of 1.00 mL/min. The inlet septum purge flow was maintained at 3 mL/min. The inlet was used in splitless mode with a purge flow of 100 mL/min delayed to start 30 seconds after injection, giving a total flow of 101 mL/min. Runs were performed in a gas saver mode with a flow of 20 mL/min set to start a minute after injection. Front inlet temperature was set at 250 °C for the entire run.

The primary oven temperature was held at 70 °C for 1 min and the temperature was ramped at 10 °C/min until 315 °C and held for 2 minutes. The secondary oven temperature and the modulator temperature offsets were 5 °C and 15 °C above the main oven respectively. A one-minute equilibration time was set for the ovens. The modulation program is listed in Table 11. The transfer line temperature was maintained at 320 °C for the entire run.

**Table 11: Modulation Timing**

| # | Start (s) | End (s) | Modulation period (s) | Hot pulse time (s) | Cool time between stages (s) |
|---|---|---|---|---|---|
| 1 | Start | 392 | 6.00 | 1.00 | 2.00 |
| 2 | 392 | End of Run | 6.00 | 1.50 | 1.50 |

## A2. MS Method

A Leco Pegasus 4D time of flight mass spectrometer (TOF-MS) with electron impact ionization was used for mass analysis. Filaments were turned off for the initial 230 seconds to delay mass acquisition until after the solvent peak. The mass scanning range was from 50 to 500 u with an acquisition rate of 200 spectra per second. The detector voltage was set at 100 V above the optimized voltage with an electron energy of -70 V. Manual mass defect mode was used with the mass defect 0 mu/ 100 u. The ion source temperature was required to reach 220 °C before starting mass acquisition.

# REFERENCES


(1) Bhalla, N., P. Jolly, N. Formisano, and P. Estrela, *Introduction to biosensors.* Essays Biochem, 2016. **60**(1): p. 1-8.

(2) Lim, J.W., D. Ha, J. Lee, S.K. Lee, and T. Kim, *Review of Micro/Nanotechnologies for Microbial Biosensors.* Frontiers in Bioengineering and Biotechnology, 2015. **3**(61).

(3) Gruhl, F.J., B.E. Rapp, and K. Länge, *Biosensors for Diagnostic Applications*, in *Molecular Diagnostics*, H. Seitz and S. Schumacher, Editors. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 115-148.

(4) Su, L., W. Jia, C. Hou, and Y. Lei, *Microbial biosensors: A review.* Biosensors and Bioelectronics, 2011. **26**(5): p. 1788-1799.

(5) Gui, Q., T. Lawson, S. Shan, L. Yan, and Y. Liu, *The Application of Whole Cell-Based Biosensors for Use in Environmental Analysis and in Medical Diagnostics.* Sensors (Basel), 2017. **17**(7): p. 1623.

(6) Khalil, A.S. and J.J. Collins, *Synthetic biology: applications come of age.* Nature Reviews Genetics, 2010. **11**(5): p. 367-379.

(7) Liu, D., T. Evans, and F. Zhang, *Applications and advances of metabolite biosensors for metabolic engineering.* Metabolic Engineering, 2015. **31**: p. 35-43.

(8) Kumar, R.R. and S. Prasad, *Metabolic engineering of bacteria.* Indian J Microbiol, 2011. **51**(3): p. 403-409.

(9) Yadav, V.G., M. De Mey, C.G. Lim, P.K. Ajikumar, and G. Stephanopoulos, *The future of metabolic engineering and synthetic biology: towards a systematic practice.* Metabolic engineering, 2012. **14**(3): p. 233-241.

(10) Alper, H.S. and J.L. Avalos, *Metabolic pathway engineering.* Synth Syst Biotechnol, 2018. **3**(1): p. 1-2.

(11) Lopreside, A., X. Wan, E. Michelini, A. Roda, and B. Wang, *Comprehensive Profiling of Diverse Genetic Reporters with Application to Whole-Cell and Cell-Free Biosensors.* Analytical chemistry, 2019. **91**(23): p. 15284-15292.

(12) Berepiki, A., R. Kent, L.F.M. Machado, and N. Dixon, *Development of High-Performance Whole Cell Biosensors Aided by Statistical Modeling.* ACS synthetic biology, 2020. **9**(3): p. 576-589.

(13) Fujimoto, H., M. Wakabayashi, H. Yamashiro, I. Maeda, K. Isoda, M. Kondoh, M. Kawase, H. Miyasaka, and K. Yagi, *Whole-cell arsenite biosensor using*

*photosynthetic bacterium Rhodovulum sulfidophilum. Rhodovulum sulfidophilum as an arsenite biosensor.* Appl Microbiol Biotechnol, 2006. **73**(2): p. 332-8.

(14) Ahmed, A., J.V. Rushworth, N.A. Hirst, and P.A. Millner, *Biosensors for whole-cell bacterial detection.* Clinical microbiology reviews, 2014. **27**(3): p. 631-46.

(15) Kylilis, N., P. Riangrungroj, H.-E. Lai, V. Salema, L.Á. Fernández, G.-B.V. Stan, P.S. Freemont, and K.M. Polizzi, *Whole-Cell Biosensor with Tunable Limit of Detection Enables Low-Cost Agglutination Assays for Medical Diagnostic Applications.* ACS Sensors, 2019. **4**(2): p. 370-378.

(16) Olaniran, A.O., L. Hiralal, and B. Pillay, *Whole-cell bacterial biosensors for rapid and effective monitoring of heavy metals and inorganic pollutants in wastewater.* Journal of Environmental Monitoring, 2011. **13**(10): p. 2914-2920.

(17) Gupta, S., M. Saxena, N. Saini, Mahmooduzzafar, R. Kumar, and A. Kumar, *An Effective Strategy for a Whole-Cell Biosensor Based on Putative Effector Interaction Site of the Regulatory DmpR Protein.* PLOS ONE, 2012. **7**(8): p. e43527.

(18) Chou, C.P., *Engineering cell physiology to enhance recombinant protein production in Escherichia coli.* Applied Microbiology and Biotechnology, 2007. **76**(3): p. 521-532.

(19) Lee, K.-H. and D.-M. Kim, *In Vitro Use of Cellular Synthetic Machinery for Biosensing Applications.* Front Pharmacol, 2019. **10**: p. 1166-1166.

(20) Jung, J.K., K.K. Alam, M.S. Verosloff, D.A. Capdevila, M. Desmau, P.R. Clauer, J.W. Lee, P.Q. Nguyen, P.A. Pastén, S.J. Matiasek, J.-F. Gaillard, D.P. Giedroc, J.J. Collins, and J.B. Lucks, *Cell-free biosensors for rapid detection of water contaminants.* Nature Biotechnology, 2020. **38**(12): p. 1451-1459.

(21) Voyvodic, P.L. and J. Bonnet, *Cell-free biosensors for biomedical applications.* Current Opinion in Biomedical Engineering, 2020. **13**: p. 9-15.

(22) Silverman, A.D., U. Akova, K.K. Alam, M.C. Jewett, and J.B. Lucks, *Design and Optimization of a Cell-Free Atrazine Biosensor.* ACS synthetic biology, 2020. **9**(3): p. 671-677.

(23) McNerney, M., C. Michel, K. Kishore, J. Standeven, and M. Styczynski, *Dynamic and tunable metabolite control for robust minimal-equipment assessment of serum zinc.* Nature Communications, 2019. **10**: p. 5514.

(24) Khambhati, K., G. Bhattacharjee, N. Gohil, D. Braddick, V. Kulkarni, and V. Singh, *Exploring the Potential of Cell-Free Protein Synthesis for Extending the Abilities of Biological Systems.* Frontiers in Bioengineering and Biotechnology, 2019. **7**(248).

(25) Perez, J.G., J.C. Stark, and M.C. Jewett, *Cell-Free Synthetic Biology: Engineering Beyond the Cell.* Cold Spring Harb Perspect Biol, 2016. **8**(12): p. a023853.

(26) Dondapati, S.K., M. Stech, A. Zemella, and S. Kubick, *Cell-Free Protein Synthesis: A Promising Option for Future Drug Development.* BioDrugs, 2020. **34**(3): p. 327-348.

(27) Ohashi, H., T. Kanamori, Y. Shimizu, and T. Ueda, *A highly controllable reconstituted cell-free system--a breakthrough in protein synthesis research.* Current pharmaceutical biotechnology, 2010. **11**(3): p. 267-71.

(28) Shimizu, Y., A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, and T. Ueda, *Cell-free translation reconstituted with purified components.* Nature Biotechnology, 2001. **19**(8): p. 751-755.

(29) Shimizu, Y., T. Kanamori, and T. Ueda, *Protein synthesis by pure translation systems.* Methods, 2005. **36**(3): p. 299-304.

(30) Carlson, E.D., R. Gan, C.E. Hodgman, and M.C. Jewett, *Cell-free protein synthesis: applications come of age.* Biotechnology advances, 2012. **30**(5): p. 1185-94.

(31) Hodgman, C.E. and M.C. Jewett, *Cell-free synthetic biology: thinking outside the cell.* Metab Eng, 2012. **14**(3): p. 261-9.

(32) Whittaker, J.W., *Cell-free protein synthesis: the state of the art.* Biotechnol Lett, 2013. **35**(2): p. 143-152.

(33) Jiang, L., J. Zhao, J. Lian, and Z. Xu, *Cell-free protein synthesis enabled rapid prototyping for metabolic engineering and synthetic biology.* Synth Syst Biotechnol, 2018. **3**(2): p. 90-96.

(34) Zemella, A., L. Thoring, C. Hoffmeister, and S. Kubick, *Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems.* Chembiochem, 2015. **16**(17): p. 2420-2431.

(35) Kim, D.-M. and J.R. Swartz, *Prolonging cell-free protein synthesis with a novel ATP regeneration system.* Biotechnology and Bioengineering, 1999. **66**(3): p. 180-188.

(36) Kim, D.-M. and J.R. Swartz, *Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis.* Biotechnology and Bioengineering, 2001. **74**(4): p. 309-316.

(37) Calhoun, K.A. and J.R. Swartz, *Energizing cell-free protein synthesis with glucose metabolism.* Biotechnology and Bioengineering, 2005. **90**(5): p. 606-613.

(38) Cole, S.D., A.E. Miklos, A.C. Chiao, Z.Z. Sun, and M.W. Lux, *Methodologies for preparation of prokaryotic extracts for cell-free expression systems.* Synth Syst Biotechnol, 2020. **5**(4): p. 252-267.

(39) Silverman, A.D., N. Kelley-Loughnane, J.B. Lucks, and M.C. Jewett, *Deconstructing Cell-Free Extract Preparation for in Vitro Activation of Transcriptional Genetic Circuitry.* ACS Synthetic Biology, 2019. **8**(2): p. 403-414.

(40) Sun, Z.Z., C.A. Hayes, J. Shin, F. Caschera, R.M. Murray, and V. Noireaux, *Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology.* J Vis Exp, 2013(79): p. e50762-e50762.

(41) Kwon, Y.-C. and M.C. Jewett, *High-throughput preparation methods of crude extract for robust cell-free protein synthesis.* Scientific Reports, 2015. **5**(1): p. 8663.

(42) Fujiwara, K. and N. Doi, *Biochemical Preparation of Cell Extract for Cell-Free Protein Synthesis without Physical Disruption.* PLoS One, 2016. **11**(4): p. e0154614.

(43) Miguez, A., M. McNerney, and M. Styczynski, *Metabolic Profiling of Escherichia coli -Based Cell-Free Expression Systems for Process Optimization.* Industrial & Engineering Chemistry Research, 2019. **58**.

(44) Wang, Y. and Y.H.P. Zhang, *Cell-free protein synthesis energized by slowly-metabolized maltodextrin.* BMC Biotechnol, 2009. **9**: p. 58-58.

(45) Pardee, K., Alexander A. Green, T. Ferrante, D.E. Cameron, A. DaleyKeyser, P. Yin, and James J. Collins, *Paper-Based Synthetic Gene Networks.* Cell, 2014. **159**(4): p. 940-954.

(46) Pardee, K., A.A. Green, M.K. Takahashi, D. Braff, G. Lambert, J.W. Lee, T. Ferrante, D. Ma, N. Donghia, M. Fan, N.M. Daringer, I. Bosch, D.M. Dudley, D.H. O'Connor, L. Gehrke, and J.J. Collins, *Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components.* Cell, 2016. **165**(5): p. 1255-1266.

(47) Byun, J.Y., K.H. Lee, Y.B. Shin, and D.M. Kim, *Cascading Amplification of Immunoassay Signal by Cell-Free Expression of Firefly Luciferase from Detection Antibody-Conjugated DNA in an Escherichia coli Extract.* ACS Sens, 2019. **4**(1): p. 93-99.

(48) Cole, S.D., K. Beabout, K.B. Turner, Z.K. Smith, V.L. Funk, S.V. Harbaugh, A.T. Liem, P.A. Roth, B.A. Geier, P.A. Emanuel, S.A. Walper, J.L. Chávez, and M.W. Lux, *Quantification of Interlaboratory Cell-Free Protein Synthesis Variability.* ACS synthetic biology, 2019. **8**(9): p. 2080-2091.

(49) Strychalski EA, R.E., *CELL-FREE (Comparable Engineered Living Lysates for Research Education and Entrepreneurship) Workshop Report.* 2020.

(50) Shin, J. and V. Noireaux, *An E. coli Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells.* ACS Synthetic Biology, 2012. **1**(1): p. 29-41.

(51) Jewett, M.C. and J.R. Swartz, *Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis.* Biotechnol Bioeng, 2004. **86**(1): p. 19-26.

(52) McNerney, M.P., F. Piorino, C.L. Michel, and M.P. Styczynski, *Active Analyte Import Improves the Dynamic Range and Sensitivity of a Vitamin B12 Biosensor.* ACS synthetic biology, 2020. **9**(2): p. 402-411.

(53) Watstein, D.M. and M.P. Styczynski, *Development of a Pigment-Based Whole-Cell Zinc Biosensor for Human Serum.* ACS synthetic biology, 2018. **7**(1): p. 267-275.

(54) Chubukov, V., A. Mukhopadhyay, C.J. Petzold, J.D. Keasling, and H.G. Martín, *Synthetic and systems biology for microbial production of commodity chemicals.* npj Systems Biology and Applications, 2016. **2**(1): p. 16009.

(55) Choi, K.R., W.D. Jang, D. Yang, J.S. Cho, D. Park, and S.Y. Lee, *Systems Metabolic Engineering Strategies: Integrating Systems and Synthetic Biology with Metabolic Engineering.* Trends in Biotechnology, 2019. **37**(8): p. 817-837.

(56) Del Valle, I., E.M. Fulk, P. Kalvapalle, J.J. Silberg, C.A. Masiello, and L.B. Stadler, *Translating New Synthetic Biology Advances for Biosensing Into the Earth and Environmental Sciences.* Front Microbiol, 2021. **11**(3513).

(57) Lowe, R., N. Shirley, M. Bleackley, S. Dolan, and T. Shafee, *Transcriptomics technologies.* PLoS computational biology, 2017. **13**(5): p. e1005457-e1005457.

(58) Riekeberg, E. and R. Powers, *New frontiers in metabolomics: from measurement to insight.* F1000Research, 2017. **6**: p. 1148-1148.

(59) He, L., Y. Xiao, N. Gebreselassie, F. Zhang, M.R. Antoniewicz, Y.J. Tang, and L. Peng, *Central metabolic responses to the overproduction of fatty acids in Escherichia coli based on 13C-metabolic flux analysis.* Biotechnology and Bioengineering, 2014. **111**(3): p. 575-585.

(60) Ohtake, T., S. Pontrelli, W.A. Laviña, J.C. Liao, S.P. Putri, and E. Fukusaki, *Metabolomics-driven approach to solving a CoA imbalance for improved 1-butanol production in Escherichia coli.* Metabolic Engineering, 2017. **41**: p. 135-143.

(61) Laviña, W.A., S.S.M. Sakurai, S. Pontrelli, S.P. Putri, and E. Fukusaki, *Metabolomics Analysis Reveals Global Metabolic Changes in the Evolved E. coli*

*Strain with Improved Growth and 1-Butanol Production in Minimal Medium.* Metabolites, 2020. **10**(5): p. 192.

(62) Yang, C., X. Gao, Y. Jiang, B. Sun, F. Gao, and S. Yang, *Synergy between methylerythritol phosphate pathway and mevalonate pathway for isoprene production in Escherichia coli.* Metabolic Engineering, 2016. **37**: p. 79-91.

(63) Wada, K., Y. Toya, S. Banno, K. Yoshikawa, F. Matsuda, and H. Shimizu, *13C-metabolic flux analysis for mevalonate-producing strain of Escherichia coli.* Journal of Bioscience and Bioengineering, 2017. **123**(2): p. 177-182.

(64) Foshag, D., E. Henrich, E. Hiller, M. Schäfer, C. Kerger, A. Burger-Kentischer, I. Diaz-Moreno, S.M. García-Mauriño, V. Dötsch, S. Rupp, and F. Bernhard, *The E. coli S30 lysate proteome: A prototype for cell-free protein production.* New Biotechnology, 2018. **40**: p. 245-260.

(65) Garenne, D., C.L. Beisel, and V. Noireaux, *Characterization of the all-E. coli transcription-translation system myTXTL by mass spectrometry.* Rapid Communications in Mass Spectrometry, 2019. **33**(11): p. 1036-1048.

(66) Garcia, D.C., B.P. Mohr, J.T. Dovgan, G.B. Hurst, R.F. Standaert, and M.J. Doktycz, *Elucidating the potential of crude cell extracts for producing pyruvate from glucose.* Synthetic Biology, 2018. **3**(1).

(67) Vilkhovoy, M., D. Dai, S. Vadhin, A. Adhikari, and J.D. Varner, *Absolute Quantification of Cell-Free Protein Synthesis Metabolism by Reversed-Phase Liquid Chromatography-Mass Spectrometry.* J Vis Exp, 2019(152).

(68) Levine, M.Z., B. So, A.C. Mullin, R. Fanter, K. Dillard, K.R. Watts, M.R. La Frano, and J.P. Oza, *Activation of Energy Metabolism through Growth Media Reformulation Enables a 24-Hour Workflow for Cell-Free Expression.* ACS synthetic biology, 2020. **9**(10): p. 2765-2774.

(69) Panthu, B., T. Ohlmann, J. Perrier, U. Schlattner, P. Jalinot, B. Elena-Herrmann, and G.J.P. Rautureau, *Cell-Free Protein Synthesis Enhancement from Real-Time NMR Metabolite Kinetics: Redirecting Energy Fluxes in Hybrid RRL Systems.* ACS synthetic biology, 2018. **7**(1): p. 218-226.

(70) Roberts, L.D., A.L. Souza, R.E. Gerszten, and C.B. Clish, *Targeted metabolomics.* Curr Protoc Mol Biol, 2012. **Chapter 30**: p. Unit30.2-30.2.24.

(71) Emwas, A.-H.M., *The Strengths and Weaknesses of NMR Spectroscopy and Mass Spectrometry with Particular Focus on Metabolomics Research*, in *Metabonomics: Methods and Protocols*, J.T. Bjerrum, Editor. 2015, Springer New York: New York, NY. p. 161-193.

(72) de Falco, B. and V. Lanzotti, *NMR spectroscopy and mass spectrometry in metabolomics analysis of Salvia.* Phytochemistry Reviews, 2018. **17**(5): p. 951-972.

(73) Haag, A.M., *Mass Analyzers and Mass Spectrometers*, in *Modern Proteomics – Sample Preparation, Analysis and Practical Applications*, H. Mirzaei and M. Carrasco, Editors. 2016, Springer International Publishing: Cham. p. 157-169.

(74) Gowda, G.A.N. and D. Raftery, *Quantitating Metabolites in Protein Precipitated Serum Using NMR Spectroscopy.* Analytical Chemistry, 2014. **86**(11): p. 5433-5440.

(75) Ren, J.-L., A.-H. Zhang, L. Kong, and X.-J. Wang, *Advances in mass spectrometry-based metabolomics for investigation of metabolites.* RSC Advances, 2018. **8**(40): p. 22335-22350.

(76) Dettmer, K., P.A. Aronov, and B.D. Hammock, *Mass spectrometry-based metabolomics.* Mass Spectrometry Reviews, 2007. **26**(1): p. 51-78.

(77) Wang, Y., Q. Chen, D.L. Norwood, and J. McCaffrey, *RECENT DEVELOPMENT IN THE APPLICATIONS OF COMPREHENSIVE TWO-DIMENSIONAL GAS CHROMATOGRAPH.* Journal of Liquid Chromatography & Related Technologies, 2010. **33**(9-12): p. 1082-1115.

(78) Smith, M., A. Miguez, and M. Styczynski, *Gas Chromatography–Mass Spectrometry Microbial Metabolomics for Applications in Strain Optimization: Methods and Protocols*. 2019. p. 179-189.

(79) Du, X. and S.H. Zeisel, *Spectral deconvolution for gas chromatography mass spectrometry-based metabolomics: current status and future perspectives.* Comput Struct Biotechnol J, 2013. **4**: p. e201301013-e201301013.

(80) Saccenti, E., H.C.J. Hoefsloot, A.K. Smilde, J.A. Westerhuis, and M.M.W.B. Hendriks, *Reflections on univariate and multivariate analysis of metabolomics data.* Metabolomics, 2014. **10**(3): p. 361-374.

(81) Miguez, A.M., M.P. McNerney, and M.P. Styczynski, *Metabolomics Analysis of the Toxic Effects of the Production of Lycopene and Its Precursors.* Front Microbiol, 2018. **9**: p. 760-760.

(82) Dromms, R. and M. Styczynski, *Systematic Applications of Metabolomics in Metabolic Engineering.* Metabolites, 2012. **2**(4): p. 1090.

(83) Su, A.M. and M.P. Styczynski, *Manipulation of metabolism in complex eukaryotic systems to control cellular state.* Current Opinion in Chemical Engineering, 2015. **10**: p. 63-69.

(84) McNerney, M.P. and M.P. Styczynski, *Small molecule signaling, regulation, and potential applications in cellular therapeutics.* Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 2017: p. e1405-n/a.

(85) Hasunuma, T., T. Sanda, R. Yamada, K. Yoshimura, J. Ishii, and A. Kondo, *Metabolic pathway engineering based on metabolomics confers acetic and formic acid tolerance to a recombinant xylose-fermenting strain of Saccharomyces cerevisiae.* Microbial Cell Factories, 2011. **10**(1): p. 2.

(86) Gold, N.D., C.M. Gowen, F.-X. Lussier, S.C. Cautha, R. Mahadevan, and V.J.J. Martin, *Metabolic engineering of a tyrosine-overproducing yeast platform using targeted metabolomics.* Microbial Cell Factories, 2015. **14**(1): p. 73.

(87) Ito, Y., T. Hirasawa, and H. Shimizu, *Metabolic engineering of Saccharomyces cerevisiae to improve succinic acid production based on metabolic profiling.* Bioscience, Biotechnology, and Biochemistry, 2014. **78**(1): p. 151-159.

(88) Watstein, D.M. and M. Styczynski, *Development of a pigment-based whole-cell zinc biosensor for human serum.* ACS Synthetic Biology, 2017. **in revision**.

(89) McNerney, M.P. and M.P. Styczynski, *Precise control of lycopene production to enable a fast-responding, minimal-equipment biosensor.* Metabolic Engineering, 2017. **43**(Part A): p. 46-53.

(90) Yoon, S.-H., Y.-M. Lee, J.-E. Kim, S.-H. Lee, J.-H. Lee, J.-Y. Kim, K.-H. Jung, Y.-C. Shin, J.D. Keasling, and S.-W. Kim, *Enhanced lycopene production in Escherichia coli engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate.* Biotechnology and Bioengineering, 2006. **94**(6): p. 1025-1032.

(91) Farmer, W.R. and J.C. Liao, *Improving lycopene production in Escherichia coli by engineering metabolic control.* Nature Biotechnology, 2000. **18**(5): p. 533-537.

(92) Kim, S.W. and J.D. Keasling, *Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in Escherichia coli enhances lycopene production.* Biotechnol Bioeng, 2001. **72**(4): p. 408-15.

(93) Alonso-Gutierrez, J., R. Chan, T.S. Batth, P.D. Adams, J.D. Keasling, C.J. Petzold, and T.S. Lee, *Metabolic engineering of Escherichia coli for limonene and perillyl alcohol production.* Metabolic engineering, 2013. **19**: p. 33-41.

(94) Xu, F., Q.P. Yuan, and H.R. Dong, *Determination of lycopene and beta-carotene by high-performance liquid chromatography using sudan I as internal standard.* Journal of chromatography. B, Analytical technologies in the biomedical and life sciences, 2006. **838**(1): p. 44-9.

(95) Lv, X., J. Gu, F. Wang, W. Xie, M. Liu, L. Ye, and H. Yu, *Combinatorial pathway optimization in Escherichia coli by directed co-evolution of rate-limiting enzymes*

*and modular pathway engineering.* Biotechnol Bioeng, 2016. **113**(12): p. 2661-2669.

(96) Spura, J., L. Christian Reimer, P. Wieloch, K. Schreiber, S. Buchinger, and D. Schomburg, *A method for enzyme quenching in microbial metabolome analysis successfully applied to gram-positive and gram-negative bacteria and yeast.* Analytical Biochemistry, 2009. **394**(2): p. 192-201.

(97) Yasid, N.A., M.D. Rolfe, J. Green, and M.P. Williamson, *Homeostasis of metabolites in Escherichia coli on transition from anaerobic to aerobic conditions and the transient secretion of pyruvate.* Royal Society Open Science, 2016. **3**(8).

(98) Faijes, M., A.E. Mars, and E.J. Smid, *Comparison of quenching and extraction methodologies for metabolome analysis of Lactobacillus plantarum.* Microbial Cell Factories, 2007. **6**(1): p. 27.

(99) Kind, T., G. Wohlgemuth, D.Y. Lee, Y. Lu, M. Palazoglu, S. Shahbaz, and O. Fiehn, *FiehnLib: Mass Spectral and Retention Index Libraries for Metabolomics Based on Quadrupole and Time-of-Flight Gas Chromatography/Mass Spectrometry.* Analytical Chemistry, 2009. **81**(24): p. 10038-10048.

(100) Vermeersch, K.A., L. Wang, J.F. McDonald, and M.P. Styczynski, *Distinct metabolic responses of an ovarian cancer stem cell line.* BMC Systems Biology, 2014. **8**: p. 134.

(101) Vermeersch, K.A., L. Wang, R. Mezencev, J.F. McDonald, and M.P. Styczynski, *OVCAR-3 Spheroid-Derived Cells Display Distinct Metabolic Profiles.* PLOS ONE, 2015. **10**(2): p. e0118262.

(102) Dhakshinamoorthy, S., N.-T. Dinh, J. Skolnick, and M.P. Styczynski, *Metabolomics identifies the intersection of phosphoethanolamine with menaquinone-triggered apoptosis in an in vitro model of leukemia.* Molecular BioSystems, 2015. **11**(9): p. 2406-2416.

(103) Wei, X., X. Shi, I. Koo, S. Kim, R.H. Schmidt, G.E. Arteel, W.H. Watson, C. McClain, and X. Zhang, *MetPP: a computational platform for comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry-based metabolomics.* Bioinformatics, 2013. **29**(14): p. 1786-1792.

(104) Dunn, W.B., D. Broadhurst, P. Begley, E. Zelena, S. Francis-McIntyre, N. Anderson, M. Brown, J.D. Knowles, A. Halsall, J.N. Haselden, A.W. Nicholls, I.D. Wilson, D.B. Kell, and R. Goodacre, *Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry.* Nat. Protocols, 2011. **6**(7): p. 1060-1083.

(105) Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-127.

(106) Xia, J., R. Mandal, I.V. Sinelnikov, D. Broadhurst, and D.S. Wishart, *MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis.* Nucleic Acids Research, 2012. **40**(Web Server issue): p. W127-W133.

(107) Smart, K.F., R.B.M. Aggio, J.R. Van Houtte, and S.G. Villas-Boas, *Analytical platform for metabolome analysis of microbial cells using methyl chloroformate derivatization followed by gas chromatography-mass spectrometry.* Nat. Protocols, 2010. **5**(10): p. 1709-1729.

(108) Bolten, C.J., P. Kiefer, F. Letisse, J.-C. Portais, and C. Wittmann, *Sampling for Metabolome Analysis of Microorganisms.* Analytical Chemistry, 2007. **79**(10): p. 3843-3849.

(109) Link, H., B. Anselment, and D. Weuster-Botz, *Leakage of adenylates during cold methanol/glycerol quenching of Escherichia coli.* Metabolomics, 2008. **4**(3): p. 240-247.

(110) Schädel, F., F. David, and E. Franco-Lara, *Evaluation of cell damage caused by cold sampling and quenching for metabolome analysis.* Applied Microbiology and Biotechnology, 2011. **92**(6): p. 1261-1274.

(111) Taymaz-Nikerel, H., M. de Mey, C. Ras, A. ten Pierick, R.M. Seifar, J.C. van Dam, J.J. Heijnen, and W.M. van Gulik, *Development and application of a differential method for reliable metabolome analysis in Escherichia coli.* Analytical Biochemistry, 2009. **386**(1): p. 9-19.

(112) de Jonge, L.P., R.D. Douma, J.J. Heijnen, and W.M. van Gulik, *Optimization of cold methanol quenching for quantitative metabolomics of Penicillium chrysogenum.* Metabolomics, 2012. **8**(4): p. 727-735.

(113) Styczynski, M.P., J.F. Moxley, L.V. Tong, J.L. Walther, K.L. Jensen, and G.N. Stephanopoulos, *Systematic Identification of Conserved Metabolites in GC/MS Data for Metabolomics and Biomarker Discovery.* Analytical Chemistry, 2007. **79**(3): p. 966-973.

(114) Drazic, A., E. Kutzner, J. Winter, and W. Eisenreich, *Metabolic Response of Escherichia coli upon Treatment with Hypochlorite at Sub-Lethal Concentrations.* PLoS ONE, 2015. **10**(5): p. e0125823.

(115) Jozefczuk, S., S. Klie, G. Catchpole, J. Szymanski, A. Cuadros-Inostroza, D. Steinhauser, J. Selbig, and L. Willmitzer, *Metabolomic and transcriptomic stress response of Escherichia coli.* Molecular Systems Biology, 2010. **6**: p. 364-364.

(116) Ye, Y., L. Zhang, F. Hao, J. Zhang, Y. Wang, and H. Tang, *Global Metabolomic Responses of Escherichia coli to Heat Stress.* Journal of Proteome Research, 2012. **11**(4): p. 2559-2566.

(117) Martin, V.J.J., D.J. Pitera, S.T. Withers, J.D. Newman, and J.D. Keasling, *Engineering a mevalonate pathway in Escherichia coli for production of terpenoids.* Nat Biotech, 2003. **21**(7): p. 796-802.

(118) Roe, A.J., C. O'Byrne, D. McLaggan, and I.R. Booth, *Inhibition of Escherichia coli growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity.* Microbiology, 2002. **148**(7): p. 2215-2222.

(119) Tuite, N.L., K.R. Fraser, and C.P. O'Byrne, *Homocysteine Toxicity in Escherichia coli Is Caused by a Perturbation of Branched-Chain Amino Acid Biosynthesis.* Journal of Bacteriology, 2005. **187**(13): p. 4362-4371.

(120) Zhu, F., L. Lu, S. Fu, X. Zhong, M. Hu, Z. Deng, and T. Liu, *Targeted engineering and scale up of lycopene overproduction in Escherichia coli.* Process Biochemistry, 2015. **50**(3): p. 341-346.

(121) Alper, H., K. Miyaoku, and G. Stephanopoulos, *Characterization of lycopene-overproducing E. coli strains in high cell density fermentations.* Applied Microbiology and Biotechnology, 2006. **72**(5): p. 968-974.

(122) Xu, J., X. Xu, Q. Xu, Z. Zhang, L. Jiang, and H. Huang, *Efficient production of lycopene by engineered E. coli strains harboring different types of plasmids.* Bioprocess and Biosystems Engineering, 2018.

(123) Kwon, Y.-C. and M.C. Jewett, *High-throughput preparation methods of crude extract for robust cell-free protein synthesis.* Sci. Rep., 2015. **5**: p. 8663.

(124) Kim, T.-W., J.-W. Keum, I.-S. Oh, C.-Y. Choi, C.-G. Park, and D.-M. Kim, *Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system.* Journal of Biotechnology, 2006. **126**(4): p. 554-561.

(125) Dopp, J.L. and N.F. Reuel, *Process optimization for scalable E. coli extract preparation for cell-free protein synthesis.* Biochem. Eng. J., 2018. **138**: p. 21-28.

(126) Didovyk, A., T. Tonooka, L. Tsimring, and J. Hasty, *Rapid and Scalable Preparation of Bacterial Lysates for Cell-Free Gene Expression.* ACS Synth. Biol., 2017. **6**(12): p. 2198-2208.

(127) Jewett, M.C. and J.R. Swartz, *Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis.* Biotechnol. Bioeng., 2004. **86**(1): p. 19-26.

(128) Vermeersch, K.A., L. Wang, J.F. McDonald, and M.P. Styczynski, *Distinct metabolic responses of an ovarian cancer stem cell line.* BMC Systems Biology, 2014. **8**(1): p. 134.

(129) Wei, X., X. Shi, I. Koo, S. Kim, R.H. Schmidt, G.E. Arteel, W.H. Watson, C. McClain, and X. Zhang, *MetPP: a computational platform for comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry-based metabolomics.* Bioinformatics, 2013. **29**(14): p. 1786-1792.

(130) Dunn, W.B., D. Broadhurst, P. Begley, E. Zelena, S. Francis-McIntyre, N. Anderson, M. Brown, J.D. Knowles, A. Halsall, J.N. Haselden, A.W. Nicholls, I.D. Wilson, D.B. Kell, R. Goodacre, and C. The Human Serum Metabolome, *Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry.* Nat. Protoc., 2011. **6**: p. 1060.

(131) Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2006. **8**(1): p. 118-127.

(132) Chong, J., O. Soufan, C. Li, I. Caraus, S. Li, G. Bourque, D.S. Wishart, and J. Xia, *MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis.* Nucleic Acids Res., 2018. **46**(W1): p. W486-W494.

(133) Leonardi, R. and S. Jackowski, *Biosynthesis of Pantothenic Acid and Coenzyme A.* EcoSal Plus, 2007. **2**(2): p. 10.1128/ecosalplus.3.6.3.4.

(134) Leonardi, R., Y.-M. Zhang, C.O. Rock, and S. Jackowski, *Coenzyme A: Back in action.* Prog. Lipid Res., 2005. **44**(2): p. 125-153.

(135) Patel, D.P., K.W. Krausz, C. Xie, D. Beyoğlu, F.J. Gonzalez, and J.R. Idle, *Metabolic profiling by gas chromatography-mass spectrometry of energy metabolism in high-fat diet-fed obese mice.* PLoS One, 2017. **12**(5): p. e0177953.

(136) Yoshida, M., K. Kashiwagi, A. Shigemasa, S. Taniguchi, K. Yamamoto, H. Makinoshima, A. Ishihama, and K. Igarashi, *A Unifying Model for the Role of Polyamines in Bacterial Cell Growth, the Polyamine Modulon.* J. Biol. Chem., 2004. **279**(44): p. 46008-46013.

(137) Fujiwara, K., T. Sawamura, T. Niwa, T. Deyama, S.-I.M. Nomura, H. Taguchi, and N. Doi, *In vitro transcription-translation using bacterial genome as a template to reconstitute intracellular profile.* Nucleic Acids Res., 2017. **45**(19): p. 11449-11458.

(138) Shi, X., T. Wu, C. M. Cole, N. K. Devaraj, and S. Joseph, *Optimization of ClpXP activity and protein synthesis in an E. coli extract-based cell-free expression system.* Sci. Rep., 2018. **8**(1): p. 3488.

(139) Miguez, A.M., M.P. McNerney, and M.P. Styczynski, *Metabolomics Analysis of the Toxic Effects of the Production of Lycopene and Its Precursors.* Front. Microbiol., 2018. **9**(760).

(140) Sikora, M. and H. Jakubowski, *Homocysteine editing and growth inhibition in Escherichia coli.* Microbiology, 2009. **155**(6): p. 1858-1865.

(141) Richards, G.R., M.V. Patel, C.R. Lloyd, and C.K. Vanderpool, *Depletion of glycolytic intermediates plays a key role in glucose-phosphate stress in Escherichia coli.* J. Bacteriol., 2013. **195**(21): p. 4816-4825.

(142) Chong, J., D.S. Wishart, and J. Xia, *Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis.* Current protocols in bioinformatics, 2019. **68**(1): p. e86.

(143) Haug, K., K. Cochrane, V.C. Nainala, M. Williams, J. Chang, K.V. Jayaseelan, and C. O'Donovan, *MetaboLights: a resource evolving in response to the needs of its scientific community.* Nucleic Acids Research, 2019. **48**(D1): p. D440-D444.

(144) Subedi, K.P., I. Kim, J. Kim, B. Min, and C. Park, *Role of GldA in dihydroxyacetone and methylglyoxal metabolism of Escherichia coli K12.* FEMS Microbiology Letters, 2008. **279**(2): p. 180-187.

(145) Ferguson, G.P., S. Tötemeyer, M.J. MacLean, and I.R. Booth, *Methylglyoxal production in bacteria: suicide or survival?* Archives of microbiology, 1998. **170**(4): p. 209-18.

(146) Pontrelli, S., R.C.B. Fricke, S.T. Teoh, W.A. Laviña, S.P. Putri, S. Fitz-Gibbon, M. Chung, M. Pellegrini, E. Fukusaki, and J.C. Liao, *Metabolic repair through emergence of new pathways in Escherichia coli.* Nature Chemical Biology, 2018. **14**(11): p. 1005-1009.

(147) Guerra, P.R., A. Herrero-Fresno, V. Ladero, B. Redruello, T.P. dos Santos, M.R. Spiegelhauer, L. Jelsbak, and J.E. Olsen, *Putrescine biosynthesis and export genes are essential for normal growth of avian pathogenic Escherichia coli.* BMC Microbiology, 2018. **18**(1): p. 226.

(148) Schwarz, D., F. Junge, F. Durst, N. Frölich, B. Schneider, S. Reckel, S. Sobhanifar, V. Dötsch, and F. Bernhard, *Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems.* Nature Protocols, 2007. **2**(11): p. 2945-2957.

(149) Voloshin, A.M. and J.R. Swartz, *Efficient and scalable method for scaling up cell free protein synthesis in batch mode.* Biotechnol Bioeng, 2005. **91**(4): p. 516-21.

(150) Yasid, N.A., M.D. Rolfe, J. Green, and M.P. Williamson, *Homeostasis of metabolites in Escherichia coli on transition from anaerobic to aerobic conditions*

*and the transient secretion of pyruvate.* R Soc Open Sci, 2016. **3**(8): p. 160187-160187.

(151) Noh, M.H., H.G. Lim, D. Moon, S. Park, and G.Y. Jung, *Auxotrophic Selection Strategy for Improved Production of Coenzyme B(12) in Escherichia coli.* iScience, 2020. **23**(3): p. 100890.

(152) Liu, M., J. Lou, J. Gu, X.-M. Lyu, F.-Q. Wang, and D.-Z. Wei, *Increasing L-homoserine production in Escherichia coli by engineering the central metabolic pathways.* Journal of Biotechnology, 2020. **314-315**: p. 1-7.

(153) Hawkins, J.S., S. Wong, J.M. Peters, R. Almeida, and L.S. Qi, *Targeted Transcriptional Repression in Bacteria Using CRISPR Interference (CRISPRi).* Methods in molecular biology (Clifton, N.J.), 2015. **1311**: p. 349-362.

(154) Dai, Z. and J.W. Locasale, *Understanding metabolism with flux analysis: From theory to application.* Metabolic engineering, 2017. **43**(Pt B): p. 94-102.

(155) Kaval, K.G. and D.A. Garsin, *Ethanolamine Utilization in Bacteria.* mBio, 2018. **9**(1): p. e00066-18.

(156) Sedighi, M., A. Zahedi Bialvaei, M.R. Hamblin, E. Ohadi, A. Asadi, M. Halajzadeh, V. Lohrasbi, N. Mohammadzadeh, T. Amiriani, M. Krutova, A. Amini, and E. Kouhsari, *Therapeutic bacteria to combat cancer; current advances, challenges, and opportunities.* Cancer Med, 2019. **8**(6): p. 3167-3181.

(157) Gupta, S., E.E. Bram, and R. Weiss, *Genetically Programmable Pathogen Sense and Destroy.* ACS synthetic biology, 2013. **2**(12): p. 715-723.

(158) Chowdhury, S., S. Castro, C. Coker, T.E. Hinchliffe, N. Arpaia, and T. Danino, *Programmable bacteria induce durable tumor regression and systemic antitumor immunity.* Nat Med, 2019. **25**(7): p. 1057-1063.

(159) Williams, N.C. and L.A.J. O'Neill, *A Role for the Krebs Cycle Intermediate Citrate in Metabolic Reprogramming in Innate Immunity and Inflammation.* Frontiers in Immunology, 2018. **9**(141).

(160) Berg JM, T.J., Stryer L, *Biochemistry*, in *Section 17.2, Entry to the Citric Acid Cycle and Metabolism Through It Are Controlled.*, W.H. Freeman, Editor. 2002: New York.

(161) Gasmi, A., M. Peana, M. Arshad, M. Butnariu, A. Menzel, and G. Bjørklund, *Krebs cycle: activators, inhibitors and their roles in the modulation of carcinogenesis.* Archives of Toxicology, 2021.

(162) Farewell, A. and F.C. Neidhardt, *Effect of Temperature on In Vivo Protein Synthetic Capacity in Escherichia coli.* J Bacteriol, 1998. **180**(17): p. 4704-4710.

(163) Treitz, C., B. Enjalbert, J.-C. Portais, F. Letisse, and A. Tholey, *Differential quantitative proteome analysis of Escherichia coli grown on acetate versus glucose.* PROTEOMICS, 2016. **16**(21): p. 2742-2746.

(164) Pletnev, P., I. Osterman, P. Sergiev, A. Bogdanov, and O. Dontsova, *Survival guide: Escherichia coli in the stationary phase.* Acta Naturae, 2015. **7**(4): p. 22-33.