NORTHWESTERN UNIVERSITY

A Cell-Free Framework for Rapid Enzymatic Pathway Prototyping and Discovery

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Chemical Engineering

By

Ashty Stephen Karim

EVANSTON, ILLINOIS

September 2018

Abstract

It is pertinent now more than ever that we find sustainable alternatives to produce chemicals. For decades, scientists and engineers have turned to biological systems to help meet societal needs in energy, medicine, materials, and more—especially when chemical synthesis is untenable. Often, biologically-produced small molecules are insufficient for production at the optimal titer, rate, or yield because natural sources are difficult to optimize and are simply not scalable (e.g., plants grow slowly). Thus, engineers seek to design enzymatic reaction schemes in microorganisms to meet manufacturing criteria. Unfortunately, this is difficult because design-build-test (DBT) cycles—iterations of re-engineering organisms to test new sets of enzymes—are detrimentally slow due to the constraints of cell growth. As a result, a typical project today might only explore dozens of variants of an enzymatic reaction pathway. This is often insufficient to identify a commercially relevant solution, consequently requiring more DBT iterations. In this dissertation, I establish and develop an *in vitro* prototyping and rapid optimization of biosynthetic enzymes approach (termed iPROBE) to test biosynthetic pathways in cell-free systems and inform cellular metabolic engineering. In this approach, cell-free cocktails for synthesizing target small molecules are assembled in a mix-and-match fashion from crude cell lysates selectively enriched with pathway enzymes. I demonstrate that this approach can quickly study pathway enzyme ratios, tune individual enzymes in the context of a multi-step pathway, screen enzyme variants for high-performance enzymes, and discover enzyme functionalities. I develop automation techniques including liquid-handling robotics and machine learning to enhance cell-free screening experiments. Furthermore, I study the cell-free physiochemical environment of cell-free biosynthetic pathways to gain control over it, enhancing our ability to engineer pathways in cell-free systems. The rapid ability to build pathways *in vitro* using iPROBE allows us to generate large amounts of data describing pathway operation under several operating conditions. I reduce that data into a quantitative metric that combines product titer, production rate, and enzyme expression (TREE score). By

simplifying the complexity of available cell-free data to one value we can now quickly screen and rank pathways in the cell-free environment and provide useful information for cellular metabolic engineering. I demonstrate iPROBE for the production of 3-hydroxybutyrate and n-butanol in *Clostridium*, an industrially relevant non-model organism. I anticipate that iPROBE will facilitate efforts to define, manipulate, and understand metabolic pathways for accelerated DBT cycles in the cell-free environment before engineering organisms. In addition, I have applied iPROBE to the study of natural product pathways often encrypted in large biosynthetic gene clusters with hopes of understanding and discovering new natural products. There are many firsts in this body of work and it is with great hopes that I share this work with the scientific community.

Acknowledgments

You can design and create and build the most wonderful place in the world. But it takes people to make the dream a reality.

- Walt Disney

The PhD is a journey into the unknown. Without the people who crossed my path along the way (many included here) I would never have made it to where I am today. I am lucky to have had inspirational teachers who guided me in my pursuit of knowledge. My elementary school teachers, Peggy Noonan and Elizabeth Schneider, taught me the joys of discovery. My high school science teachers, Joan Gambrell and Donita Hoffpauir, sparked my curiosity for chemistry and biology. My music teachers, Ed Stein, Doc Rainey, and Anna Carney brought out my creativity and perseverance. These are some of the people that paved the road to this PhD.

I have built and developed tools and frameworks providing new foundations for biochemical engineering during my PhD journey with the support of amazing people. At the heart of this journey has been the best professor I could have asked to study with: Michael Jewett. It has been my great fortune to have him as an advisor, teacher, and mentor. Mike is the quintessential professor. He is one of my biggest inspirations, and I would not be the scientist I am today without him. I am also grateful that I came to Northwestern well-equipped to create my PhD experience because of both Kate Curran, the graduate student I worked with at the University of Texas at Austin, and Hal Alper, my undergraduate advisor and mentor. Of course, my projects turned into meaningful contributions to my field with the help of my thesis committee as well. I thank Linda Broadbelt, Neil Kelleher, and Keith Tyo for their continual support through this journey.

Now imagine having to spend ~12,740 hours with the same group of people. They better be some of the best people you've ever met. Thankfully, for me, my lab-mates made this experience one-of-a-kind. Starting on day one until the day he graduated, Quentin Dudley has been a terrific mentor and colleague. As a team, we have made great contributions to our field. I'd also like to thank Jessica Perez for being a fantastic desk-mate, friend, and confidante. From hip-hop dance class to pottery to outreach and more, we spent a lot of quality time together. I am incredibly grateful to her. We also initiated the 'Texas' office with Do Soon Kim, whom I must also thank for contributing to my wonderful graduate experience. We spent both of our undergraduate and graduate research careers together in the same labs and I could not be more appreciative for the impact our conversations had on me. In addition, I thank Eric Hodgman, Jennifer Schoborg, Jian Li, Dubby Wiseman, and Antje Krüger for making my office experience intellectually-stimulating, enthusiastically fun, and one that I will dearly miss. To Jessica Stark: we joined the lab at the same time and competed for all of the same awards and I couldn't imagine a better colleague, competitor, and friend to share this journey with. To Anne D'Aquino: I appreciate the time we've spent sharing the lab's party planning committee responsibilities and being good friends. To Erik Carlson: You have been a crucial role model and friend, and I look forward to continued friendship. Of course, my tremendous thanks go out to Jake Heggestad, Sam Crowe, and Will Grubbe for being the best research mentees I could have asked for. I would also like to thank the rest of my lab-mates including but not limited to Jazzy Hershewe, Lauren Clark, Adam Silverman, Andrew Hunt, Weston Kightlinger, Adam Hockenberry, Rey Martin, Filippo Caschera, Javin Oza, and Seok Hoon Hong for the meaningful interactions that have shaped my perspective and this dissertation.

Sanity is the key to staying vigilant during any journey, so I must thank those that have kept me sane. To my board game group: Brian Larsen, Jeremy Herzberg, Jason Pike, and Alan Onnen. There are tangible benefits from playing board games regularly, but it's these people that make it truly worthwhile. To Alysse and Rey Ordóñez: playing video games with you over the years has often been the highlight of my night, allowing me to decompress from hard days in lab. There are so many others as well that have been exceptional friends and to whom I am forever grateful.

Lastly, family are those that let us go our own ways but never leave our side. I thank my dad, Faraydon Karim for always making me think I was a king who could do anything I believed in and instilling

a love for math, science, and philosophy. I thank my mom, Sharon Karim, for teaching me how to live, love, laugh, and be happy. She was always my number one fan, and I am forever thankful. I am happy we've been able to share this journey. In addition, my sister Pari Karim would like to think that I am completing my PhD because of how she led me to accept my offer at Northwestern. Whether that is true or not, she absolutely helped make it a smoother ride. To Haval and Shuan Karim for their love and support throughout my life, and to Mariam and Hero Karim for teaching me more about life than they could possibly know. To my chosen brother, Shadid Askar: for venturing on this journey with me. From picking classes for senior year of high school to majoring in chemical engineering at UT Austin to getting a PhD at Northwestern, we walk together through these intellectual and professional successes and I thank you for that. The journey doesn't seem so crazy coming home every day to the beautiful and most wonderful Caleb Rodriguez, my life partner. He has been my cheerleader and my best friend. I've been up and down with this PhD and he has followed me right along. I love you.

Perhaps the most valuable thing he taught me (his father) was that there is no contradiction between devotion to work and enjoyment of life and people

- Ahmed H. Zewail

Dedication

One can never repay one's debt to one's mother.

- Kurdish proverb

To my mother, Sharon Lee Karim. The most kind and curious person I have ever known. A storyteller.

When you walk through a storm
Hold your head up high
And don't be afraid of the dark

At the end of a storm
There's a golden sky
And the sweet silver song of a lark

Walk on through the wind
Walk on through the rain
Though your dreams be tossed and blown

Walk on, walk on
With hope in your heart
And you'll never walk alone

You'll never walk alone

Walk on, walk on
With hope in your heart
And you'll never walk alone

You'll never walk alone

Songwriters: Oscar Hammerstein Ii / Richard Rodgers

Table of Contents

List of Figures

run at 30ºC for 3 hrs. Glucose, CoA, and NAD⁺ were added to activate the n-butanol pathway and reactions were incubated for 24 h at 30 ºC. (d) A plasmid ratio optimization of pJL1-adhE1 vs. all other pJL1 constructs along with a test of three concentrations of T7 polymerase. For each, CFPS was run at 30 ºC for 3 h. Glucose, CoA, and NAD⁺ were added to activate the n-butanol pathway and reactions were incubated for 24 h at 30 ºC. All error bars represent standard deviations with n ≥ 3 independent reactions.

Figure 3.10. SDS-PAGE and autoradiogram for CFPS of each individual enzyme show full-length protein formation. (A) SDS-PAGE was run with two replicates of each CFPS reaction performed at standard conditions listed in the methods section and for 3 h at 30 ºC producing ¹⁴C-Leu incorporated protein corresponding to each enzyme in the n-butanol pathway. Molecular weights are listed at the bottom of the lanes. (B) An autoradiogram of the same gel showing that each enzyme is expressed in vitro as full-length proteins.

Figure 3.11. Protein SDS-PAGE gel and autoradiogram for CFPS of multiple enzymes produced full length product of each protein. SDS-PAGE was run with CFPS reactions containing one, two, three, four, and five DNA plasmids, building up the pathway. Reactions were each run at standard conditions listed in the methods section and for 3 h at 30 ºC producing ¹⁴C-Leu incorporated protein corresponding to each enzyme in the n-butanol pathway.

Figure 3.12. Quantification of individual enzymes in multiple protein CFPS reactions from autoradiogram indicates a lower yield of downstream pathway enzymes. (A) BL21(DE3) extract containing overexpressed proteins indicated in the key, different DNA plasmid compositions, and standard CFPS reagents identified in the methods section were incubated at 30 ºC for 3 h. Each sample was separated by SDS-PAGE and stained with Coomassie blue. The gel was exposed to autoradiogram showing full-length product of each enzyme resulting from the DNA combinations. (B) Using densitometry with ImageJ software, each lane from the gel in panel A was analyzed for band density to determine approximate, relative amounts of each pathway enzyme produced by CFPS. The bars represent the percent of each enzyme produced in vitro. After CFPS for 3 h at 30 ºC, each reaction was supplemented with glucose, NAD+, and CoA and incubated at 30 ºC for an additional 24 h to measure n-butanol production from each mix. Orange circles indicate n-butanol titers with error bars representing 1 s.d. with n ≥ 3.

Figure 3.13. Plasmid optimization for CFPS-ME of all pathway enzymes in vitro shows the ability to produce n-butanol. CFPS reactions were run in BL21(DE3) extract containing no overexpressed proteins. DNA plasmids encoding each heterologous enzyme were added in equal ratios with pJL1-adhE1 modulated as per the divisions in the circular graphics in the figure key. In the key, black represents equal amounts of plasmids encoding AtoB, Hbd, Crt, and Ter, and red represents the amount of pJL1-adhE1. CFPS reagents were added and incubated at 30 ºC for 3 h. Glucose, NAD+, and CoA were added, and samples were further incubated at 30 ºC for 24 h. n-Butanol production is measured and differs with varied concentrations of added T7 polymerase. All error bars represent 1 s.d. with n ≥ 3.

Figure 3.14. Using CFPS-ME to rapidly screen pathway enzymes. (A) n-butanol production activated by CFPS of unique Ter homologs and AdhE homologs from pJL1 constructs: Ter3 (Fibrobacter succinogenes, FS), Ter4 (Flavobacterium johnsoniae, FJ), Ter5 (Spirochaeta bajacaliforniensis, SB), Ter6 (Cytophaga hutchinsonii, CH), AdhE9 (Thermosynechococcus sp. NK55a, TN), AdhE10 (Providencia burhodogranariea, PB), and AdhE13 (Serratia marcescens, SM). Ter homologs were expressed in crude lysate mixtures containing AtoB (EC), Hbd2 (CB), Crt1 (CA), and AdhE1 (CA) overexpressed, and AdhE homologs were expressed in lysates containing AtoB (EC), Hbd2 (CB), Crt1 (CA), and Ter1 (TD) overexpressed. (B) n-Butanol production activated by CFPS putative bifunctional enzymes for Hbd and Crt activity: Hbdcrt2 (Aeropyrum camini, AC), Hbdcrt3 (Pyrobaculum aerophilum, PA), Hbdcrt4 (Sulfolobus islandicus, SI), and Hbdcrt6 (Sulfolobus acidocaldarius, SA). CFPS reactions were performed from linear DNA in crude lysate mixtures containing: (1) AtoB (EC), Ter1 (TD), and AdhE1 (CA) overexpressed to test bifunctionality, (2), AtoB (EC), Crt1 (CA), Ter1 (TD), and AdhE1 (CA) overexpressed to test Hbd functionality alone, and (3) AtoB (EC), Hbd2 (CB), Ter1 (TD), and AdhE1 (CA)

List of Tables

Chapter 1: The five-year journey – an overview

We are at the very beginning of time for the human race. It is not unreasonable that we grapple with problems. But there are tens of thousands of years in the future. Our responsibility is to do what we can, learn what we can, improve the solutions, and pass them on.

- Richard Feynman

1.1 Motivation

Imagine a world where our clothes, our food, our fuels, our medicines, our everyday lives are created biologically—this future is not too far from reality. I wake up each day excited by these extraordinary possibilities at the intersection of biology and chemistry. Advances at this intersection cure diseases, produce novel materials, and develop new understandings of the world around us. My interest in biological systems started early in high school attending my first biology course when I was 14 years old. I had not yet completed a chemistry course, but I had already learned of the types of chemicals that make our living world—nucleotides, proteins, carbohydrates, and lipids. It was mid-October, and I can remember being in class during the first molecular biology lecture learning the biochemical intricacies of transcription and translation of DNA to RNA to proteins. I remember thinking how amazed I was that we could actually see it happening—in schematic diagrams of course—that scientists were visualizing these central processes. In the case of the ribosome responsible for making proteins, we can see over 50 proteins come together with 3 RNA to assemble the macromolecular complex essential to life as we know it. The incredible complexity of this still boggles my mind and captivates me. How did so many molecular components come together and perform such simple chemistries to create such complex molecules that makes life as we know it today?

This fascination, combined with my love for chemistry instilled in me by my fantastic high school chemistry teacher, led me to further explore biological systems and their potential to create the world around us.

In college, I learned so much about myself and my interest in science. Specifically, my tenure at The University of Texas at Austin taught me three key insights that have guided throughout my career: (1) biology is like an onion in which each year we pull back new knowledge layer by layer and thus is poised to have one of the greatest capacities to change the world; (2) although ninety percent of research is not successful or fruitful, a ten percent success rate is more than enough to cause big changes in the scientific community; and (3) the field of biotechnology is in dire need of innovative and creative scientists to address the imperfections in biochemical production and the unfulfilled promises of synthetic biology. With these insights in mind, I was compelled to make my way to Northwestern to discover and share new knowledge and to contribute toward making a long-term dream of mine—creating a world in which we can make all man-made materials from biological sources—a reality.

I have no special talent. I am only passionately curious.

- Einstein

Like I said earlier, my curiosity for using biology and chemistry to engineer new worlds started early and persisted throughout my education. However, I have also carried an inexplicable love for creative endeavors and storytelling. This manifested in many ways growing up including drawing pictures, playing and even creating my own video games, and developing websites for fun. More recently, my creative outlet has involved pottery and traditional art like sequential comics. Throughout my graduate school journey, I have realized the importance of storytelling in science. The way we communicate science with other scientists and with our communities is in fact storytelling. What is a discovery that is not shared with others? Our scientific endeavors should be selfless. We should share our stories, our science. Only then can our ideas permeate and create new solutions to the toughest problems. So here is my story. This is my journey into the unknown with the chemical engineering toolbox in hand, biology and chemistry at the ready, and a whimsical spirit guiding me. What follows here is the narrative of a five-year journey into developing

powerful, enabling technologies to rationally and rapidly manipulate biological systems to produce chemicals.

Storytelling is the most powerful way to put ideas into the world today.

- Robert McKee

1.2 Dissertation Roadmap

Advances in biotechnology have allowed engineers to design biological systems for the production of chemicals, materials, and pharmaceuticals. Unfortunately, current approaches remain challenging. Major challenges include an incomplete knowledge of how cells work, unwieldy biological complexity, and the fact that it is difficult to balance intracellular fluxes to optimally satisfy active synthetic pathways while the machinery of the cell is functioning to maintain reproductive viability. New tools and approaches are needed to address these challenges and harness the synthetic and functional capabilities of biological systems. *In vitro*, or cell-free, biological systems have emerged as one approach to decouple the cell's physiological and evolutionary objectives from the engineer's objectives.

Over the past five years I've worked to develop cell-free systems for the study, discovery, and application of biosynthetic pathways. The following chapters detail the research performed towards and knowledge gained from the development of a novel cell-free framework for biosynthetic pathway prototyping and discovery. Each chapter represents a complete story that has either been published in a peer-reviewed scientific journal, is a compilation of multiple publications, or is intended for submission to such journals in the future. To tie everything together, each chapter, with the exception of Chapters 1 and 9, contains a paragraph at the beginning and end to provide its context within the dissertation.

Chapter 2 establishes a primer on cell-free systems for small molecule synthesis. This is a composite of two reviews written early in my PhD surveying the literature for examples of cell-free systems used to study and engineer biosynthetic pathways. I comment on the state-of-the-art cell-free systems for small molecule metabolite production and pathway optimization along with their challenges. This primer

and perspective provide the basis for all the developments in this dissertation. The publications this chapter is based on have also garnered wide interest by the scientific community.

Chapter 3 details the initial development of our cell-free framework for pathway prototyping and represents a key contribution from my PhD. This work showed for the first time that cell-free protein synthesis can reconstitute entire biosynthetic pathways (CFPS-ME). Without re-engineering DNA or microbes, we can build new metabolic pathways. The chapter focusses on how *in vitro* systems provide greater control and higher resolution in studying metabolism and how CFPS-ME enables the *in vitro* study of pathways in context of native metabolism. The use of linear DNA with CFPS-ME shows we can bypass *in vivo* cloning steps in pathway construction decreasing the time for pathway construction from weeks/months to hours/days. This work is the first development of cell-free protein synthesis for enzymatic pathway construction, has spawned a growing interest in cell-free metabolic systems, and gives rise to all future chapters of this dissertation.

Chapter 4 provides protocol-level detail of how to run cell-free reactions for biosynthetic pathway prototyping. Those looking to pursue cell-free systems for assembling enzymatic pathways will find this chapter most helpful. Here, I provide instructions to prepare lysates, to mix-and-match enzymes to create full pathways, and to use cell-free protein synthesis to assemble pathways. Disseminating this level of information to the community through this chapter was important to me because these cell-free approaches become more useful when others use them.

Chapter 5 deviates slightly from the original focus of my work in order to develop automation for cell-free systems. This chapter explores using liquid-handling robotics, an in vitro ribosome synthesis, assembly, and translation (iSAT) system, and machine-learning tools to optimize protein synthesis in a new strain of *Escherichia coli*. The findings here provide precedent for using liquid-handling robotics in Chapter 6 and generally for cell-free systems. The results also show the capabilities of machine-learning techniques further explored in Chapter 7.

Chapter 6 documents physiochemical optimizations of cell-free pathway prototyping systems. In reality this chapter is the true transition from Chapter 3, the initial framework, into Chapter 7, the apex application; without it, Chapter 7 would not be possible. This chapter also is my first to share with two outstanding undergraduate researchers—a true pleasure and highlight of my journey. Their contributions

make this a very special chapter for me. The key findings from this work was that an all-in-one cell-free system, with its many benefits, might not be best for controllable and reliable pathway prototyping because there are contradictory optimal physiochemical conditions for enzyme synthesis and for enzyme utilization.

Chapter 7 is the capstone work of my PhD and represents the climax of this dissertation. This chapter documents the most controllable cell-free pathway prototyping set-up to date which we term *in vitro* prototyping and rapid optimization of biosynthetic enzymes (iPROBE). We showcase its use in testing the production of 3-hydroxybutuyrate and *n*-butanol and implementing the best pathways in living organisms. This chapter shows the most extensive bridge between cell-free experimentation and cellular metabolic engineering. We test over 250 pathway combinations aided by neural-network-based machine learning algorithms. This work is a collaboration with LanzaTech, a biotechnology company located in Skokie, IL, and has been able to improve *in vivo* production of these molecules in industrially-relevant, non-model hosts, specifically in *Clostridium*. This is the most comprehensive use of cell-free systems to aid metabolic engineering efforts to date.

Chapter 8 describes my extensive foray into using the cell-free platform for natural product discovery. Medicines are often derived from fungal and bacterial natural products. Many of these molecules are catalyzed by large enzymatic complexes encoded by biosynthetic gene clusters (BGCs), which are often cryptic because we do not understand what pathways they encode. Studying BGCs is challenging because (1) native fungi and bacteria can be difficult to culture in the lab and (2) expression in model organisms can result in seemingly inactive clusters. In this chapter, I document my attempts at addressing these challenges by expanding my cell-free framework to prototype BGCs from gut bacteria, fungi*, and E. coli*, to study biosynthesis of known natural products, and, importantly, discover new ones.

Chapter 9 concludes this dissertation and my five-year journey with a summary of my accomplishments and discussion of future directions.

Be known for your science.

- Hal Alper

Chapter 2: Cell-free systems for small molecule biosynthesis – a primer

If I have seen further, it is by standing on the shoulders of giants.

- Isaac Newton

Over the last 20 years cell-free systems experienced a technical renaissance with a renewed interest to compliment current biotechnologies and develop new paradigms. This chapter surveys the state-of-the-art of cell-free systems specifically for building biosynthetic pathways. Cell-free systems have been around for over a century in their use to study the living world—from being used as a tool to study ethanol metabolism to discovering the fundamental building blocks of life. However, it has only been in the last few decades that they have garnered interest to complement cellular metabolic engineering efforts. The text in this chapter largely comes from a review published as a part of a Wiley series on biotechnology with parts selected from another review we published in Biotechnology Journal. This chapter serves an introduction to the types of cell-free systems built and developed in my dissertation and is meant to establish the foundation for what exists for cell-free metabolic engineering.

2.1 Introduction

Essentially all of today's chemicals and materials are produced from the same petroleum-based precursors: methanol, ethylene, propene, butadiene, benzene, toluene, and xylene. Unfortunately, the use of these petrochemicals has resulted in increased energy demands, shortages of feedstock, higher prices, and added climate change effects. This has motivated a growing need for greener methods of producing

fuels, pharmaceuticals, materials and commodity chemicals [1]. Biology offers one appealing approach to meet this need due to its ability to mimic the specificity and range of the commodity chemical industry's products. In fact, biologically derived substitutes for the aforementioned petroleum-based precursor molecules have already been identified [2,3].

Natural reaction cascades (*i.e.,* metabolism) in cells offer a plethora of biotransformation networks that can be manipulated to make products from simple monomer starting blocks (CO_2, sugars, etc.). Some of the oldest examples of these processes include the use of yeast for fermentation or leavening of food and alcohol. More recently, our understanding of cellular metabolism and our ability to manipulate these reactions by reading, writing, and editing DNA has vastly increased, giving rise to a growing array of chemical products (*e.g.,* 1,3-propanediol, farnesene, 1,4-butanediol, etc.[4]) that can be harnessed by redirecting cellular metabolism with metabolic engineering [5-8]. The wide range of synthetic biology tools (chassis, vectors, regulatory elements, etc.) is further expanding our ability to engineer biological systems [9,10].

While the number of metabolic engineering success stories is rapidly growing, there are limitations that currently prohibit a wide distribution of industrial-scale biotechnology. The main problem is that the process of developing and engineering living organisms to make viable production strains is expensive and time consuming [10]. In the past few decades, several notable ventures have established industrial-scale biosynthesis with metabolically engineered cells (**Table 2.1**). While the design-build-test-learn (DBTL) cycle time is decreasing, the organisms we engineer retain survival as their main objective. Because evolution has optimized the metabolism and regulation of cells for growth and adaptation, cells often resist changes that divert resources away from these objectives by altering the catalytic composition of the cell or using embedded regulatory strategies (*e.g.,* feedback inhibition of enzymes). Therefore, it is very difficult to rationally engineer intracellular fluxes to generate high titers from an active synthetic pathway while the machinery of the cell is functioning to maintain cell growth. This leads to several challenges facing the state-of-the-art. Namely, cellular fermentations provide us low titers, yields, and productivities along with an essentially limited palate of synthesizable chemicals. Many new computational, genomic, and screening tools are being developed to speed design of strains for metabolic engineering [5,11,12]. Complementary to

these efforts are new strategies to eliminate the cell's growth objectives, which can enable high yields and productivities [13,14].

Table 2.1. Examples of industrial biotechnology (time and cost).

Target	Institutions	Time to market	R&D person years
1,3- propanediol[15]	Genencor, Tate & Lyle, Du Pont	**15 years** (1992-2007)	~575
Artemisinin[10]	Amyris, UC Berkeley	**13 years** (2000-2013)	>130
Farnesene[16]	Amyris	**4 years** (2008-2012)	>250

Data from [4,10,17] and Zach Serber, Amyris Biotechnologies, DARPA Living Foundries Industry Day, personal communication

Cell-free, or *in vitro*, synthetic systems have emerged as a way of decoupling cellular growth objectives from engineering process objectives [18]. The foundational principle of cell-free synthetic systems is that we can conduct precise, complex biomolecular synthesis without using intact cells [19]. Instead, crude cell lysates or purified enzymes are used, which can be accurately monitored and modeled to provide rapid DBTL cycles. Complementing *in vivo* systems, the complexity of cell-free systems ranges from a few enzymes (2-4 enzymes) to many enzymes (more than 20 enzymes), as highlighted in several reviews [20-23]. In this chapter, we will discuss the state-of-the-art in cell-free synthetic systems with regard to cell-free metabolic engineering (CFME). A separate area of development in cell-free biology is cell-free protein synthesis (CFPS). CFPS is a useful process for understanding and harnessing biological systems. CFPS is motivating new advances in: high-throughput protein production [19,24-26], clinical manufacture [24], genetic circuit optimization [27], the construction of synthetic ribosomes [28], and incorporation of non-standard amino acids [29-36]. Perspectives on CFPS are described elsewhere [17,22,37-39]. In this review, we begin with a brief introduction of the technological capabilities of the CFME field and its potential benefits for pathway engineering, biomolecular 'breadboarding', and production. In the next section, we discuss up-to-date cell-free systems for small molecule metabolite production and pathway optimization. Finally, we examine challenges and opportunities, providing commentary about new directions in the field.

2.2 Background

Cell-Free biology has been used for many decades as a foundational research tool. For example, cell-free fermentations were first discovered in the late 19th century/early 20th century by Eduard Buchner in the attempt to elucidate the mechanism behind alcoholic fermentation of yeast cells [40]. This opened the door to the possibility of using individual enzymes to produce chemicals such as ethanol from reconstituted enzymes [41] and was a seminal study in establishing biochemistry as a field. Specifically, Buchner's work led to discoveries of how intricate metabolisms of different organisms function as well as opening the aperture to our ability to engineer metabolism for the targeted synthesis of biochemicals. In the 1960s, Nirenberg and Matthaei discovered that protein synthesis, which comprises a complex metabolic system of tens of reactions, could be activated and controlled in crude cell-free extracts [42]. They used this system to discover the genetic code. More recently, Swartz and colleagues demonstrated that cell-free systems could provide highly active energy and cofactor regeneration to fuel high-level (g/L) cell-free protein synthesis [43]. Collectively these advances along with other significant works [21,23,44] have motivated the development of capturing cell-free reactions to address the challenges of traditional metabolic engineering. There are two broad classes of cell-free systems typically used for synthesis of small molecules metabolites: purified enzymes and crude cell lysates.
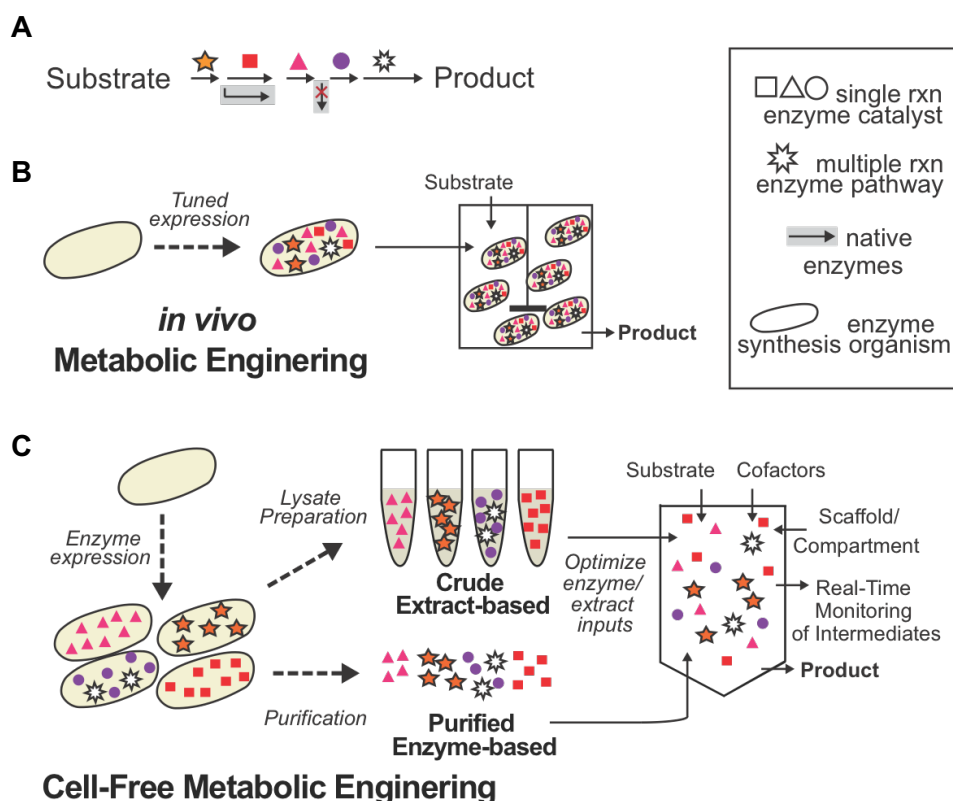
Figure 2.1. Comparison of traditional and cell-free metabolic engineering. *(A) Desired biochemical pathway. (B) Methodology for metabolic engineering in vivo. (C) Methodology for metabolic engineering in vitro.*

2.2.1   Purified enzyme systems

Purified approaches take advantage of enzymes that have been overexpressed and purified individually, then used as individual biocatalysts or recombined to assemble a pathway of interest. The benefit of these systems is that the reaction network is explicitly defined, which gives exquisite control of reaction conditions and pathway fluxes. Indeed, the clarity of the biosynthetic pathway comes from eliminating unnecessary enzymes and cellular distractions (*i.e.,* growth, other off-pathway metabolites). With respect to CFME, purified enzyme systems are used due to their distinct advantages over chemical methods, including decreased reaction time, increased product yield, greater product specificity, relatively low cost, and reduced environmental impact [45]; though, few long (>4 enzymes) pathways have been implemented. From a breadboarding perspective, cell-free systems can troubleshoot pathways difficult to achieve *in vivo* sometimes creating synthetic reaction networks leading to novel pathway design [46]. From a

production standpoint, there are already several examples of simple *in vitro* enzyme systems in industrial biocatalysis. For example, immobilized glucose isomerase has been used to harness enzymatic activity for the production of fructose [47] and the industrial production of antibiotics has been pursued with hybrid chemical and biological syntheses [48]. From these commercial examples, and others, a general rule of thumb is that purified enzyme systems are used when the product is of high value or a reaction is not possible using chemistry. Moreover, these systems are used to study the fundamental operation of biosynthetic pathways. Few industrial examples exist with synthetic enzymatic pathways (*e.g.,* more than 10 purified enzymes), in part because of the high catalyst costs (as a result of purification and stability) [43,49]. Another challenge with recapitulating pathways from purified enzymes is cofactor cost and regeneration. Despite these challenges, the majority of CFME research to date has utilized purified systems.

2.2.2   Crude cell lysate systems

As described above, as early as the 1900s scientists have looked to crude lysates as a means to harness and understand biosynthetic pathways [50]. Instead of using purified enzymes, this approach levers the extract of lysed cells as an ensemble of biocatalysts. When compared to the use of purified enzymes, crude cell lysates offer lower system catalyst costs and greater system capabilities (*e.g.,* cofactor regeneration and long-lived biocatalytic activity) [19,37]. However, crude lysates have been pursued very little for metabolic conversions *in vitro*. The presence of hundreds to thousands of active enzymes in crude lysates can make it more challenging to selectively activate and control only desired reactions. For example, the need to bypass off-pathway reactions can limit theoretical yields (an advantage of purified enzyme systems). Also, the cell-like features of these systems are partially as complex as that of whole cells. This comes with the burden of attempting to characterize the contents of the functional activities of the lysate. However, there is a growing set of examples demonstrating crude cell extracts should not be treated as a "black-box," but rather as a set of biochemical reactions that can be activated and controlled [51].

2.2.3   Comparing Biosynthesis Platforms

In comparing emerging biotechnology platforms to established ones, purified and crude extract-based CFME systems have individual merits. However, trade-offs between yield, cost, and other requirements must be carefully considered. That said, CFME systems provide a complement to *in vivo*

technologies; yet offer several distinct advantages for the design, modification, and control of biological systems (**Table 2.2**). From a biomanufacturing perspective, cell-free systems separate catalyst synthesis (cell growth) from catalyst utilization (metabolite production). This contrasts the prevailing paradigm of enclosed, cell-based, microbial "reactors." By eliminating the requirement of cell growth and diverting all carbon flux to product, CFME allows for higher theoretical yields and productivities, a wider range of products, and easier manipulation of reaction conditions. For example, cell- free production of 1,3-propanediol from glycerol (0.95 mol/mol) highlights CFME's ability to avoid by-product losses associated with traditional fermentation (0.6 mol/mol) [45]. Further, cell-free systems avoid many toxicity constraints arising from substrates, pathway intermediates, or products. It has been noted that biological processes (unlike most chemical reactions) experience heterogeneous (and often deleterious) conditions at larger scales [19]; cell-free systems may experience more chemistry-like scale-up. In one example, cell-free protein synthesis has achieved a yield and rate expansion factor of 10^6 with nearly identical performance fidelity at 100 L [40]. Finally, CFME systems may enable new opportunities in logistics and on-demand, point-of-use manufacturing.

Table 2.2. Advantages (+) and disadvantages (−) of cell-free systems.

Metric	Living cells	Cell-free systems
Pathway engineering	− Engineer's goal (overproduction) is typically opposed to microbe's goal (growth) − Endogenous regulation is difficult to predict and modify + Ability to use directed evolution	+ Easy to use chimeric enzyme pathways sourced from multiple organisms + Allow mixing with chemical catalysts/hybrid solvents that would otherwise be cytotoxic
Yield	− Carbon flux diverted to cell maintenance and byproducts	+ All carbon/energy can be directed to product
Cell wall	− Selective barrier; intercellular characterization, and product excretion can be challenging + Membrane proteins can be used	+ Direct substrate addition and product removal; easy sampling
Effect of yield toxicity	− Viability constraints (e.g. 1–2% v/v isobutanol [124])	+ No viability constraints (e.g. 12% v/v isobutanol for some pathway enzymes [49])
Stability	− Fermentation conditions affect intercellular environment	+ Reaction conditions controlled by engineer
Cost	+ Low; pending high yields, and low product separation cost	− Must incorporate cost of "catalyst" synthesis (e.g. cell growth and/or enzyme purification) − Enzyme and cofactor costs dominate
Scale-up	− Fermentation conditions are heterogeneous at industrial scale − Contamination can be catastrophic	+ Linear scale-up?
Maturity	+ Years of practical experience; well established methods	− Recently established

From a prototyping perspective, cell-free systems are well suited to rapid design-build-test cycles because they do not require the re-engineering of the entire organism with each design; merely the exogenous addition of the desired protein, cofactor, or metabolite. Consequently, there is a high degree of flexibility to model the kinetics of individual enzymes, measure metabolite fluxes in multi-step pathways, determine stability of catalysts, study the effects of redox potential on pathway performance, and experimentally isolate many other process properties that are confounded in living organisms [17]. One of the largest applications of cell-free biology to date has been in the context of pathway debugging [46].

Although cell-free technology offers many exciting advantages, challenges remain that provide opportunity for improvement. For example, most cell-free systems are not yet commercially available as production platforms. In addition, costs of in vitro systems currently exceed those of *in vivo* approaches, which limit scale-up. Despite these challenges, the benefits of CFME are inspiring new applications.

2.3 The Benefits of Cell-Free Systems

Table 2.3 summarizes the titers, yields, and productivities of recent CFME reports. These efforts have generated common rules for achieving high metabolic yields and productivities *in vitro*. Optimal metabolic conversions require, increased flux through desired pathways, decreased flux through competing reactions, and energy balancing. In short, the same concepts apply to both *in vitro* and *in vivo* metabolic engineering. Thus, development of cell-free systems should be guided by cytoplasmic mimicry to enable highly productive systems [17,43,51]. Here, we first describe advances in purified enzyme systems and follow with a discussion of crude lysate-based systems.

Table 2.3. Productivities, yields, and scales of CFME

Target Molecule	Substrate	Final Titer (mM)	Productivity (mmol L^{-1} hr^{-1})	% of Theor. Yield	Reference
		Purified Systems			
Ethanol	Glucose	28.7	1.51	57%	[23]
Isobutanol	Glucose	10.3	0.45	53%	[23]
Amorpha-4,11-diene	Mevalonate	1.66	0.28	50%	[52]
Acetyl Phosphate	F6P	~30	20.00	~100%	[46]
Hydrogen (H2)	Starch	N/A	0.44	0%	[53]
Hydrogen (H2)	Cellobiose	N/A	0.48	93%	[54]
Hydrogen (H2)	Cellopentaose	N/A	3.92	68%	[54]
Hydrogen (H2)	Xylose	N/A	1.95	95%	[55]
Xylitol (i.e. NAD(P)H)	Xylose/ Cellobiose	~18	0.38	18%	[56]
n-butanol	Glucose	3.5	0.49	82%	[57]
Lactate/lactic acid	Glucose	12.5	1.25	50%	[58]
Malate/malic acid	Glucose	5	0.625	86%	[59]
Farnesene	Acetyl-CoA	~39	0.004	35%	[44]
Isoprene	PEP	~0.5	0.13	~100%	[60]
		Crude Extract Systems			
DHAP	Glucose	9.5	19.00	83%	[61]
TDHP	Glucose	23	17.25	41%	[61]
Ethanol	Glucose	86.8	10.85	20%	[62]

2.3.1   Purified enzyme systems

Some of the first successes of cell-free systems have been in the use of purified enzymes. In 1985, Welch enabled cell-free ethanol production by reconstituting the yeast glycolytic system *in vitro* [41]. In 1992, Fessner and Walter successfully produced dihydroxyacetone phosphate (DHAP), a key metabolic intermediate in glycolysis and the Calvin cycle [63]. Fessner's system required high-energy phosphoenolpyruvate (PEP) for ATP generation and did not quantify DHAP. That said, similar to Welch's work, it demonstrated the feasibility of a multi-enzymatic pathway *in vitro*.

With this backdrop it is clear that combining enzymes to create biosynthetic pathways *in vitro* is not necessarily new, but recently there have been several success stories that utilize more complex systems of enzymes. A pioneering effort in CFME demonstrates the conversion of glucose to biofuels [23]. By developing a system with six and eight purified enzymes for ethanol and isobutanol production, respectively, Guterl and colleagues in the Sieber lab were able to show proof of effective *in vitro* metabolic engineering (**Figure 2.2A**). The reconstituted glycolytic pathway (non-phosphorylative Ennter-Doudoroff pathway) took advantage of promiscuous enzymes to downsize the number of enzymes needed from ten to four with ethanol yields at 57% and isobutanol yields at 53%. Based on the theoretical activity of individual enzymes and their concentration in the system, the theoretical maximum conversion productivity appears to be 30 mmol/L/hr; 1.5 mmol/L/hr was achieved experimentally. In the success of this pathway, an interesting question to pose is what quantity of enzymes is needed to prepare a feasible cell-free network. Furthermore, this pathway paves the way for cell-free systems to address the capabilities of using multiple cofactors and electron balancing. This example sheds light on the unexplored boundaries of cell-free systems and what they can teach us about *in vivo* biochemical pathways.

A

Glucose → Gluconate
NAD⁺ NADH

2-Keto-3-desoxygluconate

Glyceraldehyde
NAD⁺
NADH **DHAD**
Glycerate

2 Pyruvate

Acetolactate   2 Acetaldehyde
NAD⁺   NAD⁺
NADH   NADH
2,3-Dihydroxyisovalerate   **2 Ethanol**

2-Ketoisovalerate

Isobutyraldehyde NADH NAD⁺ **Isobutanol**

Guterl *et* al., 2012

B

Glucan (Gn)
H_2O

G1P

G6P

NADPH
NADP⁺
12 H_2   6 6PG   5 G6P
NADP⁺ **6 H_2O**
NADPH **6 CO_2**

6 Ru5P → *8 enzymes*

Zhang *et* al., 2007

C

G6P ← **Glucose**
ADP ATP
F6P
ATP
ADP
FBP

GAP ⇌ **TpiA** DHAP
NAD⁺   Aldehyde
NADH

1,3PG   **Product**
ATP
ADP

3PG   ATP Pyruvate
ADP   NADH
2PG → PEP   NAD⁺
            Lactate

Bujara *et* al., 2010

Figure 2.2. Biochemical pathways of key cell-free metabolic engineering achievements. *(A) Pathway from [20]. (B) Pathway from [58]. (C) Pathway from [67].*

In addition to exploring the boundaries of *in vitro* biochemical pathways, cell-free systems are also being used for rapid pathway debugging. Troubleshooting and optimizing metabolic pathways is a pernicious challenge *in vivo.* For example cases, we have to look no further than one of several success stories (*e.g.,* artemisinin production [10]) that have required more than 100 person years of development. Cell-free technologies offer some exciting opportunities for screening metabolic pathways because they provide direct access to the reaction conditions and the ability to monitor intermediate reactions. Thus, cell-free systems allow for a desirable *in vivo* pathway to be analyzed and tested for enzymatic activity *in vitro*. It is conceivable that one could take parts or modules of metabolic pathways and 'breadboard' them together to test out different pathways [27,64,65]. This circuitry approach has value for metabolic engineering and for building complex genetic circuits.

An example of *in vitro* pathway debugging was recently shown for the discovery of a synthetic pathway that mitigates the loss of carbon seen in native glycolysis [46]. Non-oxidative glycolysis (NOG) was conceived as an alternative to the standard Embden-Meyerhod-Parnas (EMP) glycolytic pathway. Using a cell-free purified enzyme system Bogorad, Lin, and Liao combined seven His-tag purified enzymes with seven purchased enzymes to create a 14-enzyme pathway that converts glucose (and xylose) to acetyl-phosphate, a metabolite easily converted to acetyl-CoA, which can enter central metabolism. Through *in*

vitro production they were able to achieve near 100% of theoretical conversion; this resulted in the ability to direct all carbon flux to acetyl-phosphate with no loss to CO_2. By establishing the *in vitro* system, expressing the NOG pathway *in vivo* required less troubleshooting of enzymes used and reaction conditions. The lack of reducing equivalents produced limits the range of products amenable to this pathway, but this exemplifies how a novel pathway can be introduced for microbial production through *in vitro* validation.

Another example demonstrating unique opportunities of the cell free approach is the synthesis of an precursor to artemisinin, amorpha-4,11-diene from mevalonate [52]. In this study, seven enzymes were His-tag purified for pathway debugging. Rather than tuning expression levels on individual enzymes *in vivo*, the cell-free approach allowed for direct modulation of pathway enzyme concentrations. Here, a design of experiments method using Taguchi orthogonal arrays explored a broad range of enzyme concentrations. Because this was done in a cell-free, isolated system, the researchers were able to determine that the fifth enzyme in the pathway, farnesyl diphosphate synthase (encoded by *IspA*), had a negative effect on production. Lowering the enzyme concentration, however, lowered the over yield of amorpha-4,11-diene. This result was actually due to precipitation of the intermediate and enzymatic product of the *IspA* enzyme, farnesyl pyrophosphate. By improving the buffer pH and salts in the cell-free environment, they produced high conversion through this multi-step pathway. While this study shows how a cell-free system can be relatively easy to manipulate, modulate, and troubleshoot, it also highlights how sometimes new challenges appear *in vitro* (such as protein precipitation) that might not be an issue *in vivo*.

Example after example demonstrates the simplicity and usefulness of *in vitro* systems for pathway optimization and achieving high theoretical yields. An instance showing near 100% of theoretical yields is the production of isoprene from phosphoenolpyruvate [60]. A pathway of 12 enzymes was implemented to achieve high yields. However, by far the largest body of work in CFME is in the production of hydrogen. Hydrogen has been pursued as an alternative energy carrier [20]. Sugar conversion to hydrogen to electricity has a higher theoretical process efficiency compared to traditional combustion of fuel and provides an opportunity to address global transportation energy challenges [66]. Recently, the Zhang group produced many reports demonstrating the production of hydrogen from sugars. Starch and water can produce high-yields of hydrogen in a cell-free environment [53]. Hydrogen can also be produced cell-free from cellobiose

[54] and xylose [55]. Also, *in silico* models of hydrogen production have been created to further expand our cell-free abilities to produce hydrogen and other chemicals [21]. Xylose has been converted to xylitol *in vitro* using oxidation of cellobiose to CO_2 as a source of energy [56]. The cell-free pathway to convert sugars to hydrogen demonstrates complexity difficult to control *in vivo* (**Figure 2.2B**). Moreover, these examples demonstrate our ability to metabolize C5 sugars that are not favored by bacteria. The Zhang group has also done tangential work to convert cellulose to starch and ethanol using hybrid purified enzyme/fermentation system; thus, building-up the toolset needed to metabolize lignocellulosic biomass [67]. While achieving high theoretical yields, the group is now seeking to improve cofactor turnover, extend peak productivities beyond eight hours, create low-cost cofactor mimics, and reduce catalyst costs to commercialize the technology [22].

Cell-free synthetic systems have also provided a platform to merge the best of synthetic chemistry and synthetic biology to produce hybrid molecules. For example, enzymes can be used to selectively make chiral bonds that would have otherwise required repeated isolation and purification of intermediates by a standard organic synthesis route. This is particularly true in the pharmaceutical industry, where biocatalysis has been shown to enable efficient and economical processes for manufacture of chiral drugs. An illustrative example is the synthesis of Stigaliptin, which is traditionally produced by asymmetric hydrogenation of an enamine at high pressure (250 psi) using a rhodium-based chiral catalyst [68]. Among other factors, the stereoselectivity limits the chemistry that can be achieved. Savile *et* al. engineered a transaminase scaffold through directed evolution to develop a biocatalyst with >99.95% e.e. for stigaliptin, a 19% reduction in total waste, and a decrease in manufacturing costs overall. In another example, Bechtold *et* al. was able to produce a precursor to the diabetes treatment, Tesaglitazar, with a 100-fold increase over what has been seen *in vivo* [69]. Also, another case shows the production of D-fagomine from glycerol and aldehyde [70].

In summary, purified enzyme systems and synthetic enzymatic pathways are opening new opportunities for high-yielding bioconversions. It is clear that cell-free synthetic systems provide benefits for validating biosynthetic pathways, building complex metabolic pathways, and expanding the chemical repertoire. While purified enzyme cascades exhibit high theoretical conversions and yields, the regulation of cofactors and stability of enzymes during reaction performance can drastically affect the efficiency of the cell-free system and is currently a limitation, which we discuss later.

2.3.2   Crude cell lysate systems

Rather than purifying individual enzymes to create a synthetic reaction network, crude cell lysates offer an alternative approach. A key difference is that native enzymes in the lysate can enable ATP and cofactor regeneration [17,71]. While this provides a benefit when compared to purified enzymes systems, the cost is added complexity. The emergence of crude extract based CFME is only new in last few years.

One of the first complex pathways activated in cell-free crude lysate systems was the conversion of glucose to dihydroxyacetone phosphate (DHAP), a key building block from the glycolysis carried out in *E. coli* extracts (**Figure 2.2C**) [61]. The ten-step pathway was activated *in vitro* utilizing natural cell metabolism, an advantage of a crude lysate system, with the triose-phosphate isomerase enzyme that utilizes DHAP removed. In other words, native enzymes present in the extract (and endogenous to *E. coli*) converted glucose to DHAP. Cofactors NAD^+ and ATP were added to direct flux through to DHAP, and similarly AMP nucleosidase was knocked out to inhibit DHAP production from ATP (independent of glucose). Of critical importance, the authors showed how to modulate cofactors to direct flux in cell-free systems and also how to adapt an *in vivo* system (*i.e.,* endogenous metabolism combined with gene knockouts) to direct an *in vitro* reaction. However, this system was limited in product yield as the byproduct lactate was generated in order to provide a sink for generated NADH.

While the overall yields were limited because of off-pathway metabolite synthesis, the system subsequently provided the basis for a real-time metabolic analysis of the cell-free reaction through mass spectrometry [72]. Real-time analysis was achieved by using a continuously stirred membrane reactor, which contained a membrane to sequester proteins while allowing small molecule exchange and be analyzed directly by ESI-MS at eight-second increments. This elegant and groundbreaking method allowed for the creation of a blueprint for enzyme regulation needed to optimize output of DHAP as well as to determine the limiting reaction in the metabolic pathway. Furthermore, a 2.5-fold increase was achieved by identification that ATP-consuming enzymes (HXK and PFK), enzymes they purified and added the system, needed close control. While this report provides highly quantitative and time-resolved capabilities to control cell-free reactions for pathway analysis, the precise genomic control needed to up-regulate HXK and PFK to appropriate levels was not demonstrated.

Beyond understanding pathway operation in a cell-like extract environment, crude lysate systems have also been exploited to create other building blocks. For example, the synthesis of triketone precursors, essential for some antibiotics, have been pursued in cell-free systems [73]. In *E. coli* crude lysates, enzymes from *B. subtilis* were overexpressed to achieve triketide lactone production. While the scientists briefly use this system to characterize enzymes in the pathway, this is a good example of how different properties such as chirality can be achieved through manipulation of enzyme activities. However, this system does not take advantage of any of the crude lysate benefits and therefore could be performed in a purified system as well. This still remains an important finding in the growing body of CFME literature.

Another key pathway that has been built in crude lysates is the production of acetyl-CoA from CO_2 and H_2 via 3-hydroxyprioponate/4-hydroxybutyrate in *P. furianosis* extracts [74]. The proposed pathway involves utilizing hyperthermophilic enzymes that can metabolize CO_2 for carbon incorporation and H_2 for redox requirements in a cycle of 13 enzymes. In each turn of the cycle two CO_2 and one acetyl-CoA yields two acetyl-CoA. Moreover, this method of carbon fixation is more carbon efficient than the Calvin cycle. To ensure flux is driven through the cycle, the authors spent efforts to characterize the first three enzymes describing the reactions from Acetyl-CoA to 3-HP. The growth temperature of *P. furianosis* is 100°C with high activity at 70°C; production at this temperature inactivates non-heterologous enzymes, yet the yields are relatively low (up to 60 mg/L 3-HP) and the cofactor regeneration system (NADPH and H_2) is still expensive.

In summary, lysate-based cell-free systems have so far seen much fewer examples than purified systems, yet the field is growing, mainly because of the potential benefits for cofactor regeneration from endogenous enzymes in the lysate. Looking forward, extracts from different organisms could be combined to create a synergistic cell-free environment for biochemical production.

2.3.3   Variations of cell-free systems

As an alternative to purified enzymes, as well as crude cell lysates, hybrid cell-free systems that straddle the divide between living and non-living also exist. For example, permeabilized cells have been explored as partial cell-free systems. This method fixes and perforates the intact membrane to maintain a reaction environment more similar to *in vivo* conditions. While there are many methods for permeabilization

of the cell membrane, most methods work by weakening the forces that hold the membrane together allowing for low-weight molecules to diffuse into the cell but large biopolymers remain [75]. These methods have been used to produce secondary metabolites. Flaviolin was produced in permeabilized *E. coli* cells with the benefit of simple reaction conditions and buffers [76]. Permeabilized cells offer benefits of the cellular environment and *in vitro* control. Also, the reusability of permeabilized cells shows that higher yields can be sustained in these systems; reactions converting fumarate to malate were repeated 6 times using the same cells showing 60 mol% yields [77]. This hybrid approach, while not best for every biosynthetic pathway, offers an interesting method worthy of further research [75].

Fuel cells are another variation on the cell-free system. The idea here is to use microbial cells produce electricity by oxidizing metabolites in the feed source; the electricity can be directly used, or the reducing power can be utilized to power further reactions. In theory this could virtually eliminate energy challenges in cell-free systems allowing for the production of high-energy molecules at low cost. However, finding the optimal feed is a challenge. Dihydrogen and dioxygen as a feed source has been used for its efficiency, but dihydrogen can be costly; methanol is one alternative that has been tried because it is more readily available [78]. Choosing the right biocatalyst to drive the fuel cell is the next key challenge. Multi-enzymatic cascades that occur naturally result in high energy densities that are beneficial to these systems; using Kreb's cycle enzymes produces a 4.6-fold power density increase over using an intact mitochondria in a pyruvate/air fuel cell [79]. Furthermore, a higher energy-density (~24µW/cm^2) pyruvate/air fuel cell was made utilizing the Kreb's cycle enzymes cross-linked together [80]. However, enzyme stability is a concern in sustaining fuel cells. One of the largest advances in cell-free fuel cells has been the switch to direct electron transfer and immobilization of enzymes for fuel cell stability [20,81]. Also, thermostability of enzymes and complete oxidation of glucose are areas that have been addressed to increase stability and maximize energy density [82-84]. In one particular example, Zhang and colleagues showed that a biosynthetic, 13-enzyme pathway in a fuel cell could produce nearly 24 electrons per glucose in maltodextrin [85]. Challenges regarding stability and energy storage are still not well characterized to produce capable fuel cells, but the ease of engineering cell-free systems over whole cells provides potential to create high-activity biofuel cells. Hybrid fuel cells that combine the best components of cell-free systems with enzyme engineering could address these challenges [20].

2.4   Challenges and Opportunities in Cell-Free Systems

With many lab-scale successes and demonstration of high theoretical yields, cell-free technologies are already showing tremendous value for pathway construction and prototyping for metabolic engineering. However, there are still many challenges to commercializing cell-free synthetic systems for biomanufacturing. Catalyst and cofactor costs remain high relative to *in vivo* counterparts. Moreover, catalyst stability is an open question particularly for synthetic enzyme pathways. In addition, the scalability on *in vitro* systems remains a concern. In recent years, much activity in the field has begun to address these challenges, which we describe below.

2.4.1   Purification

While crude lysate systems do not require individual enzyme purification [37], synthetic enzymatic pathways do. As a result, a key challenge facing those using purified enzymes is facile strategies to produce many numbers of enzymes at a low cost. The most common robust strategies used leverage affinity tags and heat purification of thermostable enzymes.

Modifying proteins can make addressing purification challenges easier and in turn controlling a cell-free system easier. Dating back to the 1980s, chemical purification such as the use of ammonium sulfate for precipitation has been used as a common technique to purify proteins and has been optimized in recent years for cell-free systems [84,86]. However, the most common strategy by far for purifying proteins today is by fusing the desired protein sequence with unique peptide sequence handles that can serve as an affinity tag (*e.g.,* FLAG-, HIS_6-, etc.). This can either be done on a plasmid, or as more recently shown by Wang *et al.*, one can use genome engineering strategies to introduce affinity tags onto the genome for purification of many enzymes at once [87]. As an added benefit, protein tags can also be used to scaffold enzymes together. Scaffolding involves binding of tagged proteins to a centralized location for temporary immobilization that can be used for purification [88]. These methods by themselves or in combination are powerful and make purification of proteins of multi-step enzymatic pathways much less daunting.

Alternatively, the use of thermostability to purify enzymes has been used to speed up the purification process. The key idea is to use enzymes from thermophilic organisms to create thermotolerant

biocatalytic modules. This is done by expressing these enzymes in *E. coli* followed by heat denaturation of all endogenous enzymes (one-step preparation). While thermophilic enzymes have been purified with this method for many years, long pathways have not been achieved until recently. Using six thermophilic enzymes, scientists were able to produce 2-deoxyriobse 5-phosphate, a nucleoside intermediate, from fructose [89]. Moreover, by combining traditional ammonium sulfate purification with thermostable enzymes optimal purifications that take advantage of both ionic strength and temperature to increase purity can be achieved. In another example, four enzymes in the glucose-3-phosphate to glucose-6-phosphate pathway were studied using heat purification for 20 minutes then a gradient of ammonium sulfate precipitation [84]. The use of heat to modulate the system has simplified and accelerated the process of purification.

Thermostable enzymes are not always best suited for desired, industrial enzymatic activities and specific activities. A key challenge for synthetic enzymatic pathways, for example, is ensuring that optimal enzyme activity for all enzymes is in the same temperature window to avoid kinetic limitations. Moreover, product inhibition has been observed. To address these issues, thermostable enzymes have been engineered. A prime example includes the characterization of two enzymes, glyceraldehyde dehydrogenase and alcohol dehydrogenase, which are thermostable [90,91]. Sieber and colleagues used directed evolution to create a thermostable glyceraldehyde dehydrogenase and alcohol dehydrogenase that were more soluble, more tolerant to substrate and products, and more efficiently utilized cofactors. These enzymes were heat purified and isolated through chromatography, providing evidence that heat purification does not have to limit the activity, stability, or strength of the enzyme.

Unfortunately, heat purification alone removes the ability of cell-free systems to use native metabolism as in a cell lysate system. To address this, chimeric pathways for natural metabolism with thermostable enzymes could be created. Ohtake and colleagues have created a thermostable reconstitution of glycolysis to make stoichiometric amounts of lactate and malate from glucose [58,92]. As seen here, a modular approach to engineering cell-free systems has created many tools to adapt *in vivo* attributes and allowed us to push the boundaries of the cell-free metabolic engineering landscape.

An example of using heat purification to develop long enzymatic pathways is the production of *n*-butanol from glucose [57]. Here, the Ohtake lab used 16 thermostable, oxygen-insensitive enzymes. Heat purification allowed them to quickly isolate their biosynthetic pathways for production of *n*-butanol. However,

stability of cofactors when using the high reaction temperatures necessary for enzyme activity is a trade-off for simple purification. Higher temperatures (>50°C) are known to degrade typical biological cofactors, which can cause issues with the use of cofactors in metabolic pathways. Some have looked at enzyme evolution to increase an enzyme's ability to use cofactors [59]. In this study, the Ohtake lab evolved malate enzymes, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and an ATP-generating phosphoglycerate kinase (PGK), to use NAD(H) instead of NADP(H). This was beneficial because NADP(H) is more thermosensitive as compared NAD(H). This further shows the power to tune enzymes within an *in vitro* system.

While there are many ways that purification strategies have developed over the last few years, there is still a lot to be done in this area to increase enzyme stability while using relatively simple and cheap purification methods.

2.4.2   Spatial organization

Cell-free systems have adapted many *in vivo* properties like organelle-like organization and molecular crowding effects to have the benefits of *in vitro* and *in vivo* synthesis [93]. An early example of this is the conversion of CO_2 to methanol in a three-enzyme pathway [94]. In this study, Obert *et* al. show the enzyme stability could be improved by running their cell-free system in an alginate-silica matrix that mimics the viscous nature of an *in vivo* cell. By achieving cellular properties in the cell-free environment, they were able to enhance production. Furthermore, compartmentalization has been utilized *in vitro* to take advantage of native benefits of localization to minimize diffusion limitations for both protein and chemical production [93]. Many *in vivo* processes utilize compartmentalization and there is a large focus in synthetic biology on engineering new ways of using this [95]. Localization of proteins during protein production and compartmentalization are two ways cellular systems take advantage of the spatial organization of biological pathways.

In cell-free systems, spatially organizing metabolic pathways can increase productivities of biosynthetic pathways by potentially sequestering pathway intermediates for diffusion limited processes, by creating high local concentrations of substrates, or stabilizing enzymes. By tethering metabolic enzymes together either by fusion proteins [96] or DNA-scaffolds [97], scientists have shown the ability to increase flux

through desired metabolic paths; this type of multi-enzymatic systems are further discussed in another review [98]. In an alternative DNA-scaffold approach, enzymes are co-localized by the addition of site-specific zinc-finger binding domains to pathway enzymes of interest. By doing so, the enzymes can be stabilized to be active for longer periods of time or also minimize diffusion limitations that reduce product formation if this is an issue [22,99,100]. In yet a third example, scaffoldin-dockerin modules have been used, mimicking *in vivo*, cellulosomes. For example, Zhang *et* al. used a cellulosome scaffold for feedstock degradation, offering both co-localization and confinement in space of enzymatic pathways for a specific design objective improving activity from 2 to several fold higher than non-complexed enzymes [101]. Broadly, the technologies that address spatial organization and co-localization in cell-free systems are expanding rapidly. This provides many avenues to pursue pathways that are diffusion-limited, pathways dependent upon a large number of reactions, parallel reaction cascades, or reactions in which the pathway enzymes can be stabilized by localization.

2.4.3   Cell-free system stability

Co-factor stability remains a central challenge for CFME. Cell-free crude-cell lysates partially address this concern by mimicking the environment of the intact cell, in one case having achieved more than 1000 turnover events of NAD^+ (which was synthesized by cells and was not added exogenously) [62]. However, there are still relatively few examples of high-level cofactor regeneration (>100 turnover events) for *in vitro* bioconversions. Without cofactor regeneration, cell-free technologies become increasing expensive due to costly additions of new cofactors. Thus, improving turnover numbers or engineering enzymes to utilize alternate cofactors is a critical hurdle for cost-effective cell-free efforts.

Cofactors, typically electron carriers and donors, are required for almost every biosynthetic pathway that utilizes energy. Cells regenerate cofactors and energy as part of complex regulation of homeostasis, but it is challenging (and not necessarily desirable) to reconstitute this system in cell-free systems. One approach is to modulate the amount of additional enzymes that regenerate cofactors to increase metabolic activity [102]. By introducing NADH oxidase, NADH dehydrogenase, and GPD1, they were able to improve the amount of NAD^+ regeneration from excess NADH. While this was performed *in vivo*, the prospect of this being done in cell-free is good. A recent study developed a robust, molecular valve to balance the

production and consumption of NADPH and NADH *in vitro* [103]. Earlier it was discussed that engineering enzymes to utilize different cofactors can increase thermostability [59]. Inherent in this is the idea that enzymes can change the cofactor that is used and ultimately change the cofactor constraints as a whole. By doing this you can adjust pathway flux for higher productivities *in vitro*. Indeed, there are a variety of works centered on exploring the ability to engineer biomimetic cofactors, which is described in several recent reviews [22,104,105].

Introducing regeneration cycles can increase the longevity of cofactors. Alternatively, using other forms of energy can improve pathway flux. By using light as an energy source, it is possible to induce NADPH regeneration. This is achieved by converting light energy through enzymes that (i) take in light, (ii) transfer electrons, and (iii) mediate electron use. One illustrative example of this type of system uses eosin Y as a photosensitizing dye, triethanolamine as an electron donor, and [Cp*Rh(bpy)H$_2$O] as an electron mediator [106]. Here the system was used to improve cytochrome P450 activity. A newer technology involves cofactor regeneration by directing cofactors toward enzymes that regenerate them via improved mixing [107]. This form of cofactor organization utilizes magnetic nanoparticles to create a Brownian motion effect and in turn a denser cell-free environment. The magnetic particles allow for cofactors to stay near enzymes after use of cofactors. Enzymes co-localized to the enzymes that used the cofactors regenerate these cofactors at a faster rate than enzymes not localized in this fashion. While the breadth of technologies that have been developed to stabilize cofactor use, there is a lot of work to be done in cofactor engineering to increase reaction lengths and final product titers.

2.4.4   Modeling

With the development of technologies for purification, spatial organization, and cofactor regeneration comes the need for more powerful computational tools. Computational approaches in evaluating metabolic engineering efforts have often benefited the design of *in vivo* systems. The immediate challenge in adapting *in silico* methods to cell free systems requires movement away from a cell growth-focused approach. Despite this challenge, new computational techniques have been developed to model cell-free systems [49]. In this study, Bujara and Panke develop network topology analysis in cell-free systems based on genome scale metabolic models (similar to flux analyses performed for *in vivo* systems). These

methods span the production of proteins, the kinetics of biosynthetic pathways, and non-growth-related analysis of reactions. Optimization of cell extracts and inputs is sometimes the limiting factor in efficient cell-free reactions. An example of a cell-free optimization approach is the *in silico* evaluation of a complex multi-enzymatic system producing H_2 from cellobiose [21]. In this study, Ardao *et* al. were able to model improved productivities from ~2 to 355 mmol/L/hr. This involved multi-objective optimizations to understand how differences in the cell-free environment interact with multiple enzymes together. Moreover, a typical modeling approach with metabolic engineering is to map metabolic pathways and their respective kinetics. Computer design of metabolic pathways *in vivo* and *in vitro* are improving production through the use of a variety of new tools that include such as 'Metabolic Tinker' for breadboarding pathways [108]. Computational tools to design DNA for pathway enzymes has also been developed [109]. Additional effort to discover, predict, and design novel conversion pathways is embodies by programs such as (BNICE) [110,111], PathPred [112], UM-PPS [113], Retropath [114], and others. The development of new tools unique for advancing CFME will grow as our understanding of the cell-free environment is better understood.

2.5 Recent Advances

Cell-free systems for metabolic engineering expand our ability to study metabolism and access biochemicals. There have been many recent advances in purified enzyme and crude lysate systems. Purified enzyme systems recently explored the production of 1,3-propanediol, hydrogen, and farnesene. The Zeng lab produced 1,3-propanediol from glycerol showing that cell-free systems can avoid loss of product to byproduct pathways [115]. The Zhang lab has continued to develop the cell-free hydrogen production platform to use additional sugars, like sucrose, and to co-utilize sugars [116,117]. Additionally, Zhu and colleagues produced farnesene from acetyl-CoA with nine purified enzymes utilizing NADPH and ATP [44]. The purified enzyme system was used to prototype enzymes to improve *in vivo* titers. These works demonstrate that purified enzyme systems continue to expand the cell-free palate of biochemicals and perform robust cell-free analyses of metabolic pathways.

Crude lysate systems have expanded both in the realm of metabolic engineering and cell-free protein synthesis metabolism. Ninh and colleagues explored one-pot conversion of glucose to pyruvate

and lactate using crude *E. coli* extracts expressing over nine thermotolerant glycolytic enzymes [118]. The Park lab has continued to develop yeast cell-free bio-ethanol production by incorporating encapsulation of cell-free systems as well as testing different simultaneous saccharification and fermentation across varying pH, temperature, and physiochemical environments [119,120]. Additionally, Kay and Jewett expressed a full heterologous pathway to 2,3-butanediol (2,3-BDO) [121]. Here, a single strain of *E. coli* was engineered to express three pathway enzymes necessary to make 2,3-BDO. With no strain optimization, Kay and Jewett observed a maximal synthesis rate of 2,3-BD of ~10 g/L/h, titer of ~80g/L, and theoretical yield of 71%. This work led to the development of an extract mixing approach to CFME where individual lysates containing individually overexpressed heterologous enzymes are mixed together to construct a biosynthetic pathway activated by the addition of simple substrates (Dudley and Jewett, in preparation) and use cell-free protein synthesis to drive metabolic conversions (Karim and Jewett, in preparation). Furthermore, energy metabolism in crude lysates has been shown to improve cell-free protein synthesis. Caschera and Noireaux activated a poly-sugar metabolism to efficiently fuel cell-free protein synthesis improving on the state-of-the-art cell-free protein synthesis platform [122]. Anderson *et al.* also explored a different energy source for cell-free protein synthesis [123]. This work showed that glucose could be used to fuel yeast cell-free protein synthesis. Crude lysate systems could transform the way we prototype biosynthesis.

2.6 Summary

Cell-free systems can be composed of purified enzymes, crude-cell lysates, or variations thereof. They provide a complement to traditional cellular systems used in biotechnology. By removing the constraints of cellular growth and decoupling catalyst synthesis from catalyst utilization, biosynthetic pathways can direct the entirety of a given substrate towards the production of a single product. This potentially provides interesting advantages for prototyping pathways and biomanufacturing. While large-scale production of multi-step enzymatic systems have yet to be demonstrated commercially, the cell-free approach is already being used for biomolecular breadboarding and pathway debugging. This is clearly evident from the examples provided in the text and highlighted in **Table 2.3**. Such advances show the potential power of CFME. That said CFME is still in its early stages, and there are several key challenges

to be addressed in the areas of purification, cofactor regeneration systems, and reconstitution of beneficial cellular mechanisms such as molecular crowding and compartmentalization. Underlying these needs is the overarching goal to lower catalyst and cofactor costs. Looking forward, recent advances in CFME suggest new opportunities. For example, cell-free systems may provide advantages for making products that can't be made *in vivo* due to toxicity or product purification limitations. They might also offer exciting new directions in the synthesis of hybrid biochemicals comprised of parts derived from organic syntheses and parts derived from biological syntheses. The ability to not only use enzymes from multiple organisms but unique metabolisms from across the phylogenic spectra are also on the horizon. In tandem with biomanufacturing prospects, cell-free systems show potential in prototyping biosynthetic pathways to guide enzyme discovery and *in vivo* production.

2.7 Conclusions

The development of cell-free tools and platforms for metabolic and pathway engineering are rapidly growing. This chapter introduced the beginnings of the field in purified and crude systems, both having distinct advantages. Challenges exist in the areas of purification, spatial organization, stability, and modelling, but not without excellent opportunities. From fuels cells to highly-active metabolic pathways, cell-free systems are paving the way for sustainable engineering design-build-test cycles and biomanufacturing. Since publication of these reviews we have published several papers developing robust cell-free capabilities to probe biosynthetic pathways highlighted in the 'Recent Advances' section of this chapter. I look forward to seeing the growing body of literature in cell-free systems for biosynthetic pathway prototyping and biomanufacturing.

Chapter 3: A novel cell free framework for rapid biosynthetic pathway prototyping and discovery

---

Anything is Possible

- Michael Jewett

This chapter represents the foundation of the dissertation. Within my first few months in the lab I took the task of outlining several projects I wanted to research based on several government agency reports on biotechnology on the horizon. This project is a result of that list. I decided I would try to produce *n*-butanol as it was a well-characterized biosynthetic pathway and it was the longest pathway anyone in the lab had tried to make in a cell-free system. As I was trying to imitate the approaches we had available in the lab, it did not make sense to me that we were boasting that cell-free systems could make building biosynthetic pathways quicker, yet we were still using live *E. coli* to produce our enzymes of interest—a process that can take weeks just to change one enzyme. At that moment I had the idea to couple a technique our lab has expertise in, cell-free protein synthesis, with our cell-free pathway prototyping approaches. It was this idea that became this chapter and was the first of its kind, shifting the paradigm of how we build pathways in cell-free systems. Rather than making enzymes *in vivo* we can now make them all *in vitro*.

3.1 Abstract

Speeding up design-build-test (DBT) cycles is a fundamental challenge facing biochemical engineering. To address this challenge, we report a new cell-free protein synthesis driven metabolic

engineering (CFPS-ME) framework for rapid biosynthetic pathway prototyping. In our framework, cell-free cocktails for synthesizing target small molecules are assembled in a mix-and-match fashion from crude cell lysates either containing selectively enriched pathway enzymes from heterologous overexpression or directly producing pathway enzymes in lysates by CFPS. As a model, we apply our approach to *n*-butanol biosynthesis showing that *Escherichia coli* lysates support a highly active 17-step CoA-dependent *n*-butanol pathway *in vitro.* The elevated degree of flexibility in the cell-free environment allows us to manipulate physiochemical conditions, access enzymatic nodes, discover new enzymes, and prototype enzyme sets with linear DNA templates to study pathway performance. We anticipate that CFPS-ME will facilitate efforts to define, manipulate, and understand metabolic pathways for accelerated DBT cycles without the need to reengineer organisms.

3.2 Introduction

For decades scientists and engineers have turned to engineering biological systems to help meet societal needs in energy, medicine, materials, and more[7,124-126]. This has been an attractive, sustainable way to produce small molecules, especially when chemical synthesis is untenable[4,127]. The ability to harness organisms that naturally produce molecules of interest has expanded the available chemical palate[128,129]. Often when natural producers are insufficient for production at the optimal titer (g l^{-1}), yield, or volumetric productivity (g l^{-1} h^{-1}), engineers seek to design biosynthetic pathways and regulatory processes in cells to meet certain manufacturing criteria[130,131]. For example, introducing heterologous pathways into model microorganisms and engineering them to maximize a particular biosynthesis has led to large scale production of 1,3-propanediol, farnesene, and artemisinin with many more on their way to market[4,17]. Efforts to make these molecules have resulted in success, but not without a great deal of challenges.

Bringing a biosynthetic molecule to market usually involves countless hours of design-build-test (DBT) cycles[13]. The production of *n*-butanol is a prime example of these challenges. A series of *Clostridia* species are natural producers of *n*-butanol during acetone-butanol-ethanol fermentation, and *Clostridia acetobutylicum* and *Clostridia beijerinckii* are two of which are commonly used in commercial *n*-

butanol plants[132]. However, these species are difficult to engineer because of a biphasic metabolism, unknown regulation patterns, and a limited number of species-specific engineering tools[133]. Heterologous expression of *Clostridia* metabolism in model microorganisms like *Escherichia coli* and *Saccharomyces cerevisiae* allows *n*-butanol production to be more easily engineered but can be accompanied by lower titers[134,135]. Starting with heterologous expression of the *n*-butanol pathway as a baseline, scientists have been able to increase titers dramatically by knocking out genes from genomes[134], increasing redox driving forces by introducing pathway-independent enzymes[136], and identifying homologous enzymes with better activities[137]. Years of iterative metabolic engineering led to these advances, but titers are still not high enough and scale-up is often too unpredictable to outcompete natural producers for commercial production[138]. As is the same for many biosynthetic pathways, we cannot quickly enough identify optimal biosynthetic systems and discover the best sets of enzymes that work together as a group. Therefore, metabolic engineering remains costly and time-consuming[5,10].

A key challenge in metabolic engineering is balancing the tug-of-war that exists between the cell's physiological and evolutionary objectives on one side and the engineer's process objectives on the other. Put another way, it is very difficult to balance intracellular fluxes to optimally satisfy a very active synthetic pathway while the machinery of the cell is functioning to maintain reproductive viability. Other challenges include: (i) the need for reliable computational selection and design of enzyme homologs for pathway design, (ii) the limited number of feasible homologs and genetic constructs that can be searched in any one project, and (iii) the unknown effects of optimal pathway enzyme expression on the entire metabolic system[109,139,140].

Many established and emerging technologies seek to address these challenges. For example, metabolic flux analysis and genome engineering offer generalized capabilities to modify living organisms for improving product titers[14,141]. In addition, coupling machine-learning algorithms to multiplexed designs can accelerate efforts to rationally engineer cells[142]. However, DBT cycle time remains a limitation[143]. *In vitro* systems offer a complementary, yet underutilized approach to speed up DBT cycles with some potential advantages[17,64,65,71]. For example, the open reaction environment allows for the addition of components such as cofactors and intermediates at any time during a cell-free reaction, which can be maintained at precise concentrations. In addition, cell-free systems have no cell viability constraints.

Furthermore, the cell-free format permits DBT iterations without the need to reengineer organisms[65], with the potential to reduce DBT cycle time[64]. Cell-free metabolic engineering (CFME), or using cell-free techniques to aid metabolic engineering efforts, is emerging as a complementary approach to existing strategies for carrying out biomolecular transformations of interest with *in vitro* ensembles of catalytic proteins, prepared from purified enzymes or crude lysates of cells[23,41,57,118,121,144-146].

In this work, we develop a cell-free protein synthesis driven metabolic engineering (CFPS-ME) framework to accelerate DBT cycles for optimizing and debugging biosynthetic pathways (**Figure 3.1A**). The foundational principle is that we can construct discrete metabolic pathways through combinatorial and modular assembly of lysates containing enzyme components produced by overexpression in the lysate chassis strain or by cell-free protein synthesis (CFPS). We focus on using CFPS because these systems can help address the growing demand for simple, inexpensive, and efficient protein production technologies for a wide array of applications[17,43,71,147-150]. In addition, processes that take days or weeks to design, prepare, and execute in cells can be done more rapidly in a cell-free system, because no time-consuming cloning steps are needed[151]. Three recent advances enable the use of CFPS for CFME. First, Jewett *et* al. demonstrated the ability to stimulate highly active energy and cofactor regeneration pathways in crude cell lysates[51]. Second, Kay and Jewett showed that crude cell lysate-based cell-free systems from *E. coli* could fuel highly active heterologous metabolic transformations[121]. Third, Dudley and Jewett established the ability to build a heterologous biosynthetic pathway by mixing lysates each containing individually overexpressed heterologous enzymes (in preparation). The mix-and-match approach has many advantages including only needing to express one enzyme in each strain, not needing to fine-tune expression, and being able to directly monitor and sample the reaction environment. Here, we extend this approach by demonstrating modular assembly of pathways through the ability to enrich lysates with biosynthetic enzymes using well-defined experimental conditions and CFPS. It is important to note that our goal in this work was not to develop cell-free systems for the highest product titer, an engineered strain for best *in vivo* synthesis of *n*-butanol, or industrial applicability. However, we do show that CFPS-ME offers an even faster approach (hours rather than days) for building pathways directly in lysates for the purpose of enzyme selection and pathway design.
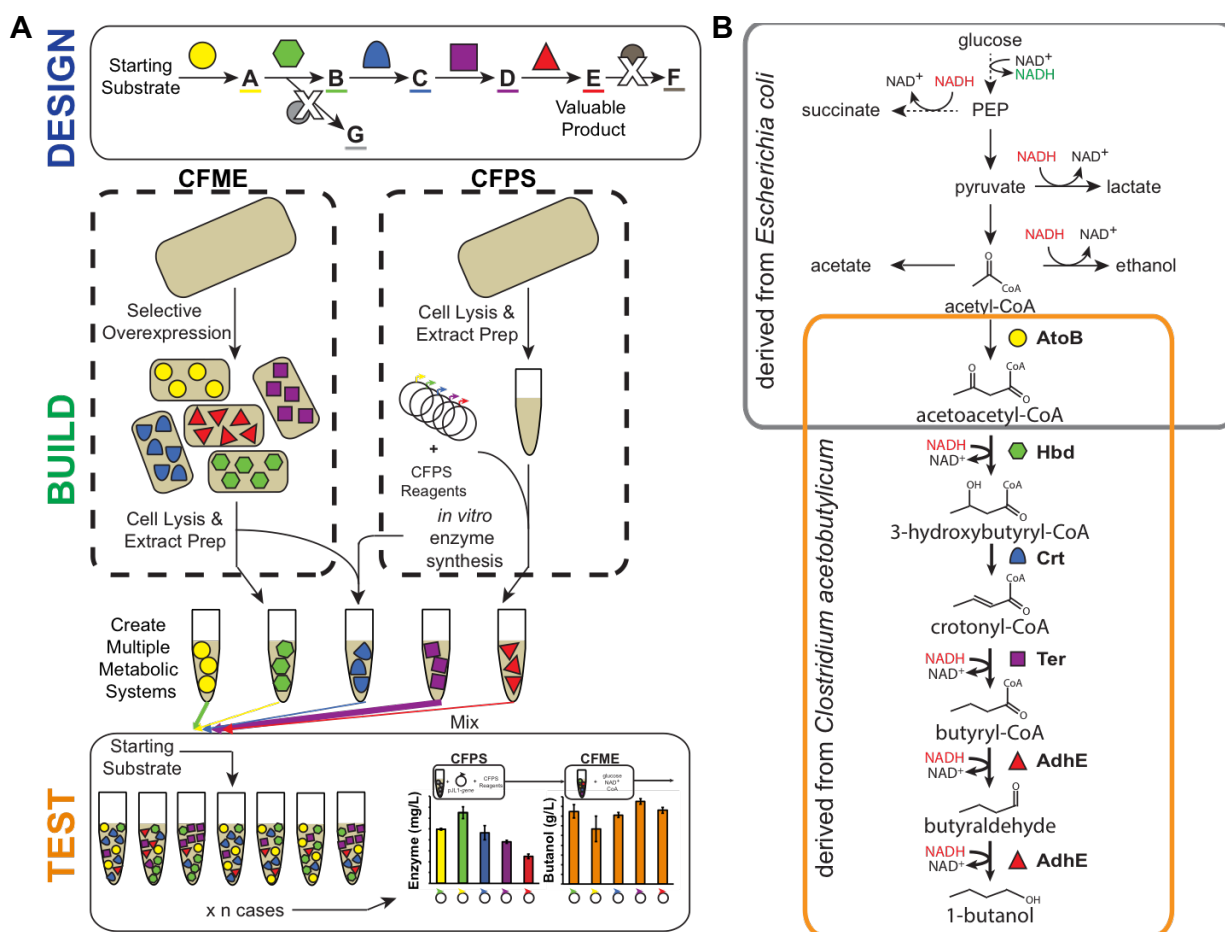
Figure 3.1. A cell-free framework for pathway prototyping demonstrated with a 17-step n-butanol model pathway. *(A) Methodology for cell-free metabolic engineering (CFME) and cell-free protein synthesis driven metabolic engineering (CFPS-ME). (B) Schematic (non-stoichiometric) representation of the constructed biosynthetic n-butanol pathway. Acetyl-CoA is generated through E. coli's natural glycolysis and funneled into the C. acetobutylicum-derived CoA-dependent pathway to produce n-butanol. The butyryl-CoA dehydrogenase (Ter) here is from Treponema denticola. Four NADH molecules are needed to produce one molecule of n-butanol.*

To demonstrate CFPS-ME, we selected the model *n*-butanol biosynthetic pathway derived from

Clostridia metabolism involving CoA intermediates (**Figure 3.1B**). Endogenous glycolytic enzymes convert

glucose to acetyl-CoA, the starting intermediate for *n*-butanol synthesis, another *E. coli* enzyme takes

acetyl-CoA to acetoacetyl-CoA, and heterologous enzymes convert acetoacetyl-CoA to *n*-butanol. We first

show the ability to mix five crude lysates each with selectively overexpressed enzymes to activate the entire

17-step *n*-butanol production pathway *in vitro* with high yield and productivities. We then establish the

CFPS-ME concept by modularly building the *n*-butanol pathway with lysates harboring heterologous

pathway enzymes expressed by CFPS or having been overexpressed in the chassis source strain. We apply this framework to rapidly screen enzymes for optimal pathway operation and enzyme discovery. We expect that the CFPS-ME framework will increase the resolution at which we can manipulate biosynthetic pathways by examining enzyme kinetics, measuring metabolic flux, determining catalyst stability, studying redox effects, and prototyping metabolism.

3.3   Materials and Methods

3.3.1   Bacterial strains and plasmids

E. coli NEB Turbo™ (NEB) was used in plasmid cloning transformations and for plasmid preparation. *E. coli* BL21(DE3) (NEB) was used for protein overexpression and for preparation of all extracts (see **Table 0.1** for strain details). A modified version of pET-22b (Novagen/EMD Millipore), used in previous studies[121], was used for all constructs for *in vivo* over-expression of proteins. For *in vitro* expression of proteins, the pJL1 vector was used. Carbenicillin (100 µg ml⁻¹) was used with the pET vector system and kanamycin (50 µg ml⁻¹) was used with the pJL1 vector system.

Gibson assembly was used for seamless construction of plasmids (see **Table 0.1** for plasmid details). Each gene and vector was amplified via PCR using forward and reverse primers designed with NEB's Gibson Assembly Designer (New England Biolabs, Ipswich MA, USA) and purchased from IDT and Phusion® High-Fidelity DNA polymerase (Finnzymes, Thermo Scientific Molecular Biology) (see **Table 0.2** for genes and enzymes and **Table 0.3** for primer details). Both PCR products were cleaned and mixed with Gibson assembly reactants and incubated at 50 °C for 60 min. Plasmid DNA from the Gibson assembly reactions were immediately transformed into *E. coli* NEB Turbo cells. Propagated constructs were purified using an EZNA Plasmid Mini Kit (Omega Bio-Tek). Completed constructs were used to transform *E. coli* BL21(DE3).

Codon optimized versions of each gene were identified using IDT's codon optimization online tool (Integrated DNA Technologies®, Coralville, USA) and NCBI's Basic Local Alignment Search Tool (National

Center for Biotechnology Information, U.S. National Library of Medicine, Bethesda MD, USA). These genes were purchased from Gen9, Inc. (Cambridge MA, USA) (see **Table 0.4** for codon-optimized sequences).

3.3.2   Cell Extract Preparation

E. coli BL21(DE3) cells (see **Table 0.1** for strains) were grown in 2 × YTPG media (16 g l^{-1} tryptone, 10 g l^{-1} yeast extract, 5 g l^{-1} NaCl, 7 g l^{-1} potassium phosphate monobasic, 3 g l^{-1} potassium phosphate dibasic, 18 g l^{-1} glucose). These cells were cultured at the 50 ml scale in 250 ml baffled tunair shake flasks (IBI Scientific, Peosta, IA) in a 37 °C incubator with vigorous shaking at 250 rpm. The cultured cells were monitored by spectrophotometry (Genesys 10S UV-Vis, Thermo Fisher Scientific, Waltham, MA). When cells reached OD_{600} = 0.6-0.8, the cultures were induced with 0.1 mM IPTG. After induction cultures were grown for 4 h at 30 ºC. Antibiotics were not used during cell growth. The cells were harvested by centrifuging at 8,000g at 4 °C for 15 min and were washed two times with cold S30 buffer (10 mM Tris-acetate (pH 8.2), 14 mM magnesium acetate, and 60 mM potassium glutamate). After final wash and centrifugation, the pelleted wet cells were weighed, flash frozen in liquid nitrogen, and stored at −80°C. The thawed cells were suspended in 0.8 ml of S30 buffer per 1 g of wet cell mass. In order to lyse cells by sonication, thawed and suspended cells were transferred into 1.5 ml microtube and placed in an ice-water bath to minimize heat damage during sonication. The cells were lysed using a Q125 Sonicator (Qsonica, Newtown, CT) with 3.175 mm diameter probe at frequency of 20 kHz and 50% of amplitude. The input energy (Joules) was monitored and 830 J was used for 1.4 ml of suspended cells. The lysate was then centrifuged twice at 21,100g at 4 °C for 15 min. All of prepared cell extract was flash frozen in liquid nitrogen and stored at −80 °C until use.

3.3.3   Extract Protein Quantification

The total protein concentration of the extracts was measured by Quick-Start Bradford protein assay kits (Bio-Rad) with a bovine serum albumin standard. The extracts were subsequently run on a Coomassie-blue stained NuPAGE Bis-Tris 12% SDS-PAGE gel with MOPS buffer (Life Technology, Grand Island, NY). The SeeBlue Plus2 pre-stained ladder (Life Technology, Grand Island, NY) was used and ~10 µg of total protein for each sample was loaded on the gel.

3.3.4   CFME Reactions

Reactions were carried out in 1.5 ml Eppendorf tubes at 37 °C in 25 µl volumes. Each reaction consisted of mixing five extracts, containing one enzyme overexpressed each, to complete the biosynthetic *n*-butanol pathway (2 mg ml^{-1}) along with magnesium glutamate (8 mM), ammonium glutamate (10 mM), potassium glutamate (134 mM), glucose (200 mM), dipotassium phosphate (10 mM, pH 7.2), Bis Tris (100 mM), NAD (1 mM), ATP (1 mM), and CoA (0.5 mM), unless otherwise noted. Reactions were terminated by adding 5%w/v trichloroacetic acid in a 1:1 ratio. Precipitated proteins were pelleted by centrifugation at 15,000g for 10 min. The supernatant was stored at −80 °C until analysis.

3.3.5   CFPS-ME Reactions

CFPS reactions were performed to express enzymes involved in *n*-butanol production prior to starting the CFME portion of the reactions using a modified PANOx-SP system[152,153]. A 25 µl CFPS reaction in a 1.5 ml microcentrifuge tube was prepared by mixing the following components: ATP (1.2 mM); GTP, UTP, and CTP (0.85 mM each); folinic acid (34.0 µg ml^{-1}); *E. coli* tRNA mixture (170.0 µg ml^{-1}); T7 RNA polymerase (100 µg ml^{-1}); 20 standard amino acids (2 mM each); nicotinamide adenine dinucleotide (NAD; 0.33 mm); coenzyme-A (0.27 mM); spermidine (1.5 mM); putrescine (1 mM); potassium glutamate (130 mM); ammonium glutamate (10 mM); magnesium glutamate (12 mM); phosphoenolpyruvate (PEP; 33 mM), and cell extract (10 mg ml^{-1}). For each reaction plasmid was added at ~13.3 or ~26.6 µg ml^{-1}. The *n*-butanol production portion of the reaction was initiated by spiking in glucose (200 mM) and additional reagents (NAD, CoA) noted throughout the manuscript.

3.3.6   Quantification of protein produced in vitro

Cell-free protein synthesis reactions were performed as noted above (Section 2.5) with radioactive ^{14}C-Leucine (10 µM) supplemented in addition to all 20 standard amino acids. We used trichloroacetic acid (TCA) to precipitate radioactive protein samples. Radioactivity of TCA-precipitated samples was measured by liquid scintillation counting to then quantify the protein produced as previously reported (MicroBeta2; PerkinElmer)[51,152]. These reactions were also run on a Coomassie-stained SDS-PAGE gel and exposed by autoradiography. Autoradiographs were imaged with a Typhoon 7000 (GE Healthcare Life Sciences,

Pittsburgh, PA). Multiple proteins produced *in vitro* were further quantified by gel image intensity comparisons using ImageJ (NIH).

3.3.7   n-Butanol Quantification

High-performance liquid chromatography (HPLC) was used to analyze the components in the reactions. *n*-Butanol was measured with an Agilent 1260 series HPLC system (Agilent, Santa Clara, CA) via a refractive index (RI) detector. Analytes were separated using the Aminex HPX-87H anion exchange column (Bio-Rad Laboratories) with a 5 mM sulfuric acid mobile phase at 55 °C and a flow rate of 0.6 ml min^{-1}. Commercial standard of *n*-butanol was used for quantification of experimental samples by linear interpolation of external standard curves. An example chromatogram for *n*-butanol is given in **Figure 3.2**.
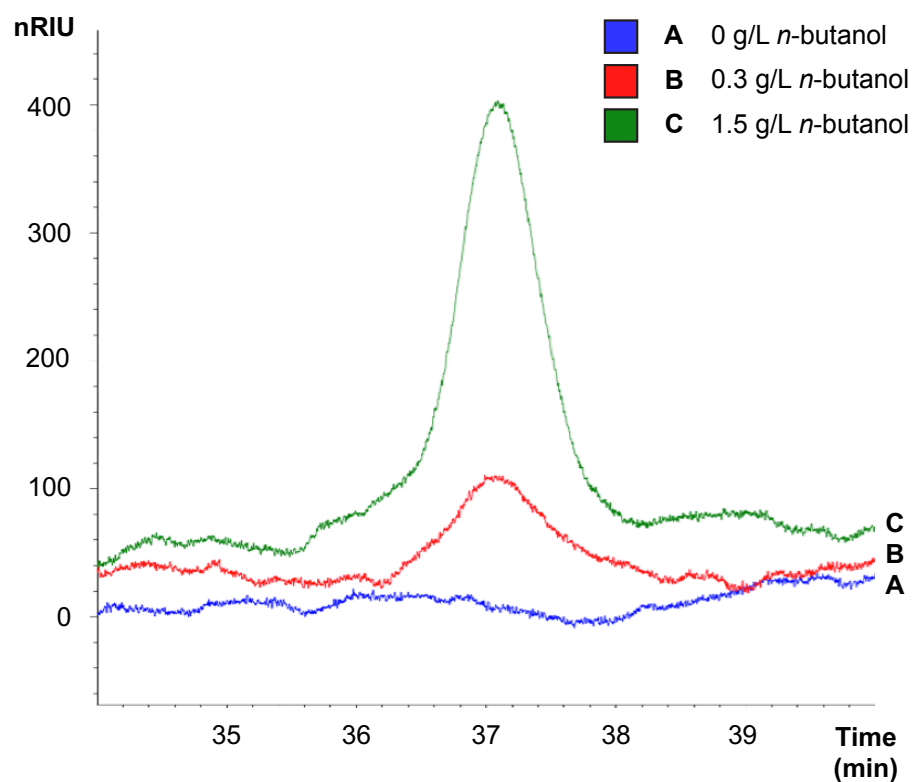
Figure 3.2. Example Chromatograms for n-butanol quantification. *Three chromatogram n-butanol peaks are overlaid with retention time on the x-axis and relative intensity units of the y-axis. Peak A represents an example cell-free reaction with no n-butanol produced from the reaction. Peak B is a cell-free reaction with ~0.3 g/L of n-butanol produced. Peak C is a cell-free reaction with ~1.5 g/L of n-butanol produced.*

3.4 Results

In developing a framework for biosynthetic pathway prototyping, we constructed a 17-step pathway for the production of *n*-butanol. *n*-butanol synthesis was selected as a model because of its importance as a potential biofuel, it is easily quantified by HPLC, and it has multiple heterologous steps. We sought to combine *E. coli*'s endogenous 11-step glycolytic pathway from glucose to acetyl-CoA (AcCoA) with the *Clostridia*-derived six-step *n*-butanol pathway from AcCoA (**Figure 3.1B**). The idea that natural energy and cofactor regeneration would be harnessed in the lysate to fuel *n*-butanol production is a distinct break from typical *in vitro* approaches, which use purified enzymes[144]. Complementary to those systems, our approach allows for studying pathway performance in a setting that better mimics the *in vivo* operation (*e.g.,* from glucose rather than AcCoA). The crude lysate system also allows us to focus on expressing only the necessary heterologous enzymes to complete the entire pathway. These enzymes include a thiolase to merge two AcCoAs followed by a number of dehydrogenases to perform a series of reductions through CoA intermediates to obtain *n*-butanol (See **Table 0.2** for Genes and Enzymes).

3.4.1   Cell-Free Metabolic Engineering for n-butanol Production

To enable cell-free biosynthesis of *n*-butanol, we first introduced genes encoding the five enzymes needed to convert AcCoA to *n*-butanol individually into our extract source strains, in this case BL21(DE3) (See **Table 0.1** for Strains and Plasmids and **Table 0.3** for Primers). We selected two homologs each for hydroxybutyryl-CoA dehydrogenase (Hbd), crotonase (Crt), and bifunctional aldehyde/alcohol dehydrogenase functionalities. For the thiolase (AtoB) and butyryl-CoA dehydrogenase (Ter) we chose *E. coli*'s endogenous enzyme and a widely used enzyme from *Treponema denticola*, respectively. Next, we selectively overexpressed each heterologous enzyme in separate strains using a tightly controlled T7 promoter and strong ribosome binding site. As expected, we observed that the heterologous proteins were overexpressed as the dominant bands, with the exception of Hbd1, on an SDS-PAGE gel (**Figure 3.3A**). The low expression of Hbd1 is likely due to RBS used for expression.
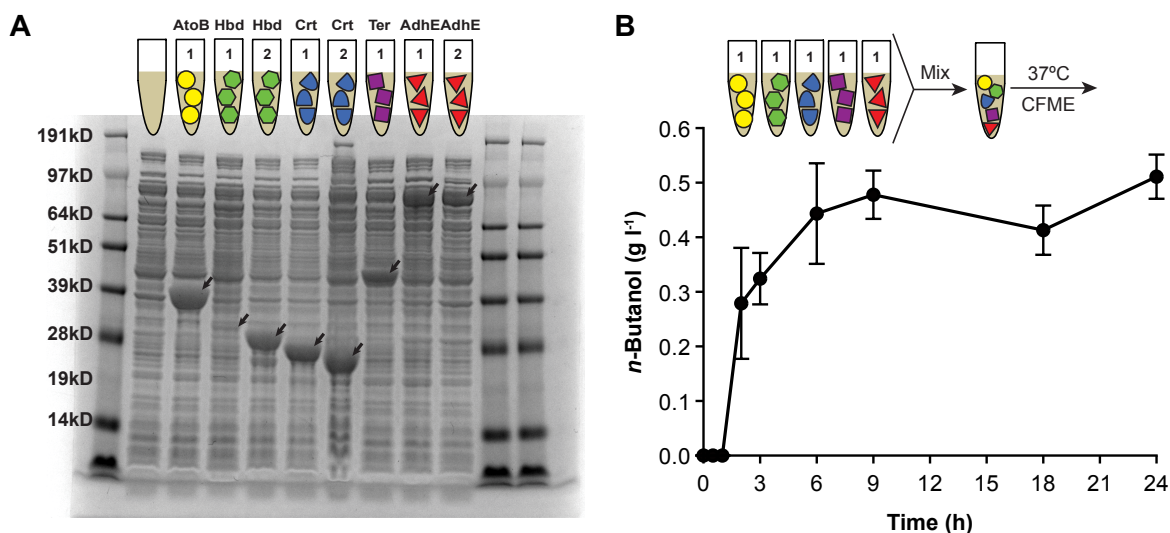
Figure 3.3. Biosynthesis of n-butanol achieved via CFME of a coupled E. coli and C. acetobutylicum metabolic pathway. *(A) Via SDS-PAGE, the gel verifies the selective overexpression of pathway enzymes in* E. coli *BL21(DE3) crude cell lysates: AtoB (*Escherichia coli*), Hbd1 (*Clostridia acetobutylicum, CA*), Hbd2 (*Clostridia beijerinckii, CB*), Crt1 (*Clostridium acetobutylicum, CA*), Crt2 (*Pseudomonas putida, PP*), Ter (*Treponema denticola, TD*), AdhE1 (*Clostridium acetobutylicum, CA*), and AdhE2 (*Clostridium pasteurianum, CP*). (B) CFME reactions for* n-*butanol production from glucose were carried out using five crude lysates mixed together (1:1:1:1:1 based on total protein quantification) with glutamate salts (Mg*$^+$*, NH*$_4$$^+$*, K*$^+$*), phosphate (K*$_2$*HPO*$_4$*), buffer (Bis Tris), and cofactors (ATP, CoA, NAD*$^+$*). These lysates individually contained AtoB (EC), Hbd1 (CA), Crt1 (CA), Ter1 (TD), and AdhE1 (CA) selectively overexpressed at 37 ºC. Error bars represent standard deviations with n ≥ 3 independent reactions.*

After lysis and extract preparation, we then reconstituted the 17-step pathway from glucose to *n*-butanol by mixing equal total protein concentrations of five separate extracts containing each enzyme. Specifically, we started with the following enzyme set: *E. coli's* AtoB*, C. acetobutylicum's* Hbd, Crt, and AdhE2, as well as Ter from *T. denticola*. This set was chosen to include most of *C. acetobutylicum's* enzyme set, one of the most widely used sets for *n*-butanol production, along with previously identified best enzymes for thiolase and butyryl-CoA dehydrogenase functions[136,138]. Upon incubation with essential substrates, salts, and cofactors (*e.g.,* magnesium, potassium, and ammonium salts, glucose, phosphate, buffer, NAD, CoA, ATP), we assessed *n*-butanol synthesis in 25 µl CFME batch reactions carried out for 24 h at 37 ºC via high performance liquid chromatography (HPLC). We observed production of 0.51 ± 0.04 g l^{-1} *n*-butanol (~0.05 mol *n*-butanol / mol glucose) over the course of a 24 h reaction (**Figure 3.3B**), without any optimization to improve titers. As expected, we also observed lactate, acetate, and ethanol as byproducts seen in previous reports of *n*-butanol production, which could be addressed through genome modifications

(*e.g.*, deletion of *ldh* gene in the source strain)[138]. Butanol production shows that both the heterologous pathway and endogenous glycolysis is activated with cofactors being regenerated. However, *n*-butanol production stops after ~9 h. In previous work, substrate depletion was shown to be the most typical cause for reaction termination[121]. One way to avoid this limitation is to run reactions in fed-batch or continuous reactor set-ups or use substrates that are metabolized slower (e.g. polymeric sugars). Except in few instances[41,60], limited cofactor regeneration has historically plagued *in vitro* synthetic enzymatic pathway conversions[23,144,146]. Here, however, native glycolytic enzymes in the lysate provide a simple route to fuel highly active heterologous metabolic conversions. For example, to produce ~7 mM *n*-butanol we would need ~56 NADH turnover events, exceeding typical turnover numbers of ~5-20 for purified *in vitro* systems[144].

Following demonstration of activating *n*-butanol synthesis, we next aimed to modularly build *n*-butanol synthesis pathways with different enzyme homologs to improve pathway performance. We cycled through multiple distinct ensembles of enzymes by mixing and matching lysates containing different versions of enzymes necessary to complete the biosynthetic *n*-butanol pathway. Trying out different homologs in this manner allowed us to quickly identify a better set of enzymes producing *n*-butanol at 0.84 ± 0.19 g l^{-1} (0.09 mol *n*-butanol / mol glucose) (**Figure 3.4A**). Specifically, we showed that Hbd2 from *C. beijerinckii* enabled a 65% increase in *n*-butanol synthesis titers over Hbd1 from *C. acetobutylicum*. A follow-up experiment doubling the Hbd1 enzyme did not alter the amount of *n*-butanol produced, suggesting that this increase was not due to discrepancies in enzyme concentrations in the lysate (**Figure 3.5**). However, further studies of these enzymes would elucidate whether the observed *n*-butanol production was a result of using BL21(DE3) extract without heterologous genes expressed (used for normalization), which may have more active glycolytic and byproduct pathways that could divert flux away from *n*-butanol.
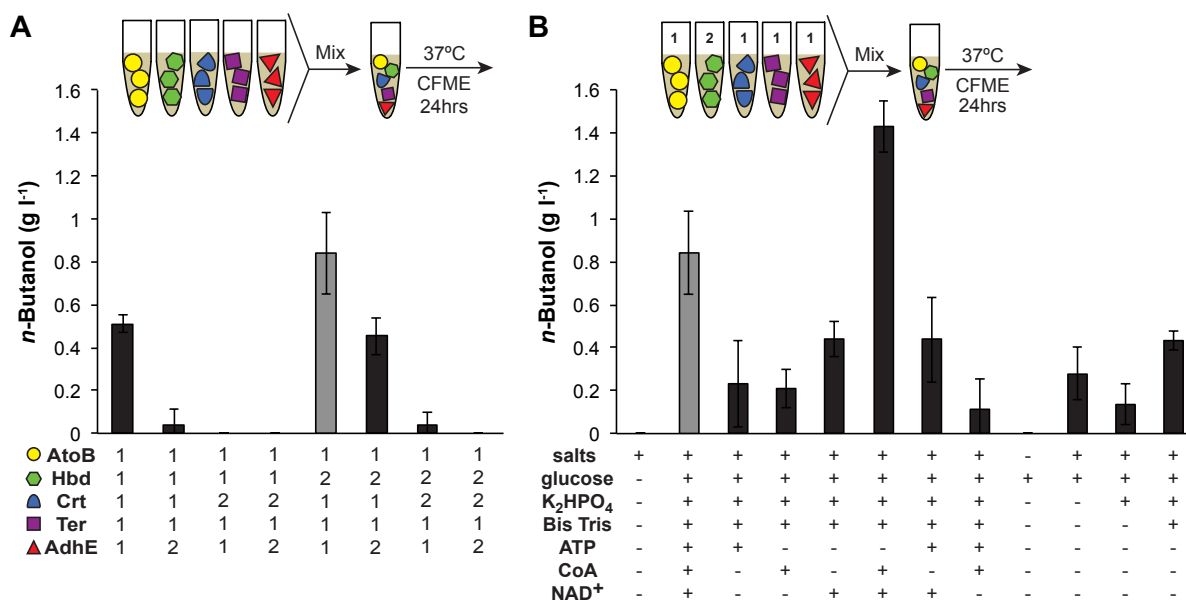
Figure 3.4. Enzyme and physiochemical optimizations lead to increased yields of CFME n-butanol production. *(A) Reactions for* n-*butanol production from glucose were performed using different sets of five crude lysates mixed together to obtain unique combinations of selectively overexpressed enzymes with AtoB, Hbd, Crt, Ter, and AdhE activities. Lysate mixes were combined with glutamate salts (Mg*$^+$*, NH*$_4^+$*, K*$^+$*), phosphate (K*$_2$*HPO*$_4$*), buffer (Bis Tris), and cofactors (ATP, CoA, NAD*$^+$*) and incubated for 24 h at 37 ºC. (B) To enhance yields and optimize pathway performance, a physiochemical optimization was performed with or without glutamate salts (Mg*$^+$*, NH*$_4^+$*, K*$^+$*), phosphate (K*$_2$*HPO*$_4$*), buffer (Bis Tris), and cofactors (ATP, CoA, NAD*$^+$*)) of cell-free reactions producing* n-*butanol. Reactions incubated for 24 h at 37 ºC. The grey bars represent the same recipe in (A) and in (B). All error bars represent standard deviations with n ≥ 3 independent reactions.*
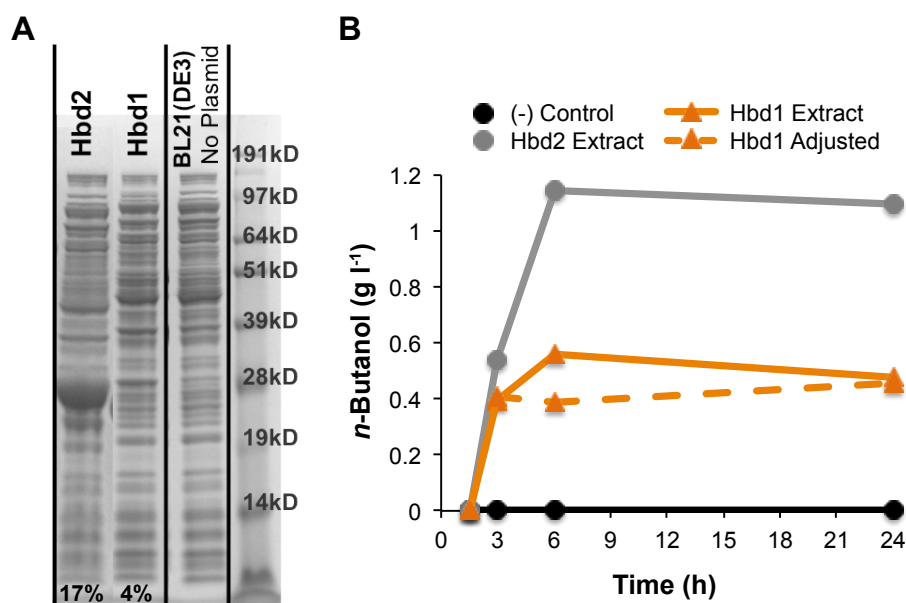
Figure 3.5. Adjusting extracts for relative concentrations of selectively overexpressed Hbd1 and Hbd2 does not affect overall n-butanol production by CFME. *(A) BL21(DE3) extract containing no overexpressed proteins, overexpressed Hbd1 extract, and overexpressed Hbd2 extract were each separated by SDS-PAGE and stained with Coomassie blue. Using densitometry with ImageJ software, each lane was analyzed for band density to determine approximate, relative amounts of overexpressed protein. Hbd1 extract contained ~4% Hbd1 protein, and Hbd2 extract contained ~17% Hbd2 protein. (B) Each Hbd extract was mixed with extracts containing the other pathway enzymes (AtoB, Crt, Ter, and AdhE) and CFME reactions were run for 24 h (n=1) to make* n-*butanol. The cases include Hbd2 extract, Hbd1 extract, and Hbd1 extract adjusted to contain approximately the same amount of overexpressed Hbd protein as the Hbd2 extract. Total extract concentration was kept constant at 10 µg ml⁻¹ using a BL21(DE3) extract containing no heterologously expressed proteins to adjust the Hbd extracts. The observed discrepancies in 'Hbd1 Extract' and 'Hbd1 Adjusted' are likely an artifact of sample size.*

While the selection of enzymes is crucial to improving *n*-butanol production, the value of each physiochemical parameter of the cell-free system also affects *n*-butanol production and becomes key in further optimization and debugging of the pathway. To demonstrate the facile nature of combinatorial optimizations in our cell-free framework, we explored changes in the ionic composition because the composition of salts added to *in vitro* systems affects the systems' performance[28,51,152,154]. Specifically, we tested the effect of using glutamate, acetate, and chloride salts on *n*-butanol production and found that glutamate salts perform more than 15% better than the other salt compositions (**Figure 3.6**). Our results are consistent with previous works, which have shown that glutamate salts better mimic the intracellular cytoplasmic conditions of *E. coli* to co-activate authentic biological processes such as the *in vitro* co-activation of central metabolism, oxidative phosphorylation, and protein synthesis[51].
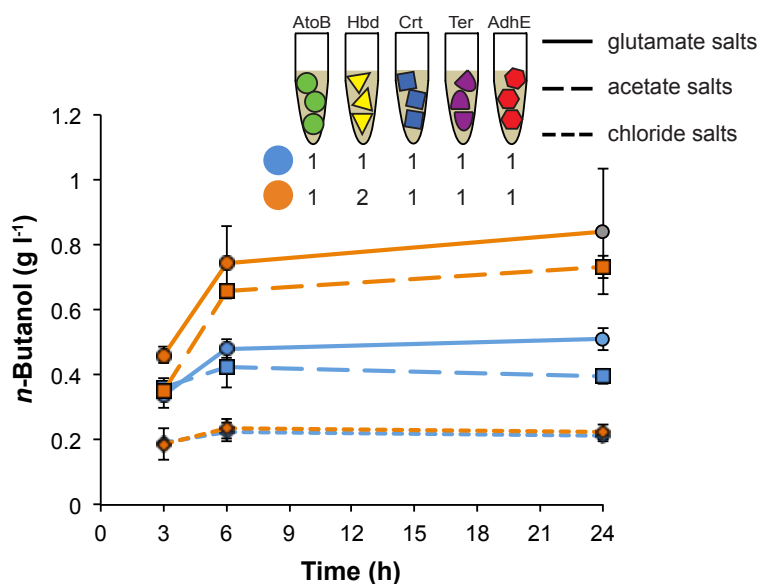
Figure 3.6. Inorganic glutamate salt solutions perform the best in CFME reactions. *Two different sets of 5 crude lysates mixed together containing selectively overexpressed enzymes with AtoB, Hbd, Crt, Ter, and AdhE activities were used to produce* n-*butanol from glucose. Blue is the original set of extracts (containing Hbd1) and orange is the best enzyme set (containing Hbd2). Lysate mixes were combined with glutamate salts (Mg^+, NH_4^+, K^+) (solid line), acetate salts (long dashed line), or chloride salts (short dashed line). To activate metabolism and start CFME reactions phosphates (K_2HPO_4), buffer (Bis Tris), and cofactors (ATP, CoA, NAD^+) were added and incubated for 24 h at 37 ºC. All error bars represent 1 s.d. with n ≥ 3.*

Beyond studying pathway performance by altering the ionic composition, the states of critical cofactors (organic molecules necessary for enzyme catalysis) can also be studied. The balance of cofactors, such as oxidized and reduced NAD, is critical to energy regeneration within the lysate by also the heterologous pathway under investigation. In our cell-free framework, the lack of a cell wall enables direct sample acquisition, reaction monitoring, and control. We used this flexibility to study the impact of the ratio of initial cofactors in the reaction to see the ratio's effect on *n*-butanol production. We found that the ratio of NAD(H) at the start of the reaction (*e.g.,* NAD:NADH: 1:0, 1:0.5, 1:1, 0.5:1, 0:1), keeping the total cofactor concentration at 0.5 mM, plays a minimal role in how much *n*-butanol can be produced (**Figure 3.7**). This suggests that metabolism in the lysate may control the overall levels of reduced and oxidized cofactor, which is consistent with data from Kay and Jewett[121].
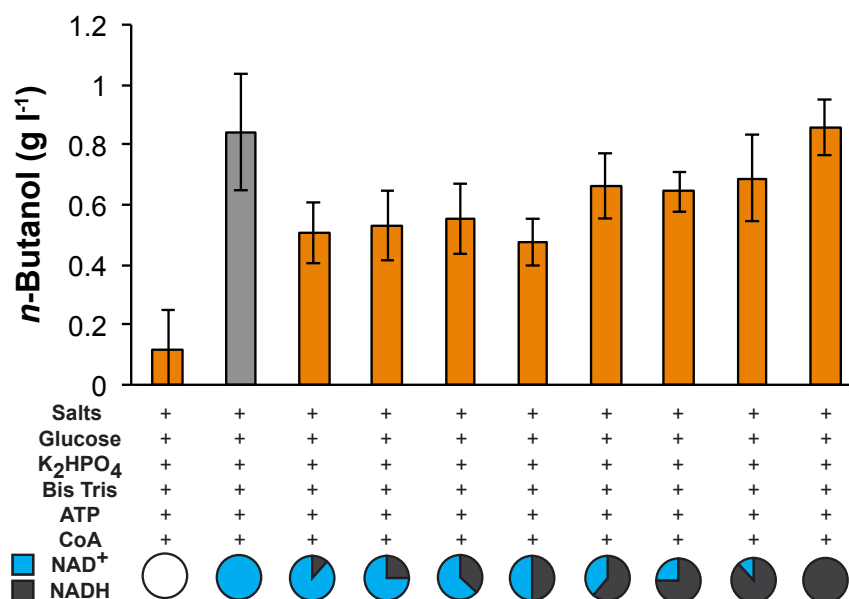
Figure 3.7. Comparing initial NAD+ and NADH ratios show relatively little effect on n-butanol production in CFME. *Reactions for* n-*butanol production from glucose using the best set of crude lysates mixed together containing selectively overexpressed enzymes with AtoB, Hbd, Crt, Ter, and AdhE activities (determined as best by previous experiments) were run for 24 h and incubated at 37 ºC. Extract mixes were combined with glutamate salts (Mg+, NH4+, K+), phosphates (K2HPO4), buffer (Bis Tris), and cofactors (ATP, CoA, NAD+, NADH) to initiate CFME reactions. In the key, plus signs (+) indicate the addition of each individual components at levels described in the methods section. The divisions in the circular graphics accompanying each bar with blue being NAD+ and black being NADH indicate ratios of NAD+ to NADH. The total NAD(H) concentration is 0.5 mM. All error bars represent 1 s.d. with n ≥ 3.*

Understanding that some components play more of a role in pathway performance than others, we next performed a number of reactions to identify which added components are necessary for *n*-butanol production with a particular interest in the three added cofactors (ATP, NAD, and CoA). The supplementation of cofactors to cell-free reactions would be costly and hinder industrial practicality of this technology if it were proposed as a biomanufacturing platform. In our study of cell-free systems as a prototyping framework, we surprisingly found that omitting ATP boosts *n*-butanol production by greater than 180% from 0.84 ± 0.19 g l^{-1} to 1.43 ± 0.12 g l^{-1} (0.11 mol *n*-butanol / mol glucose) (**Figure 3.4B**). More unexpectedly, by just adding salts to mimic the cytoplasm and glucose as a starting substrate we are able to produce *n*-butanol at 0.28 ± 0.12 g l^{-1}. In other words, if lysates are prepared without dialysis, as we have done, cofactors remaining in the lysate are sufficient for the cell-free transformation and do not need to be added. Collectively, our results here show that the cell-free framework offers a strategy to explore how

enzyme variants, substrates, cofactors, ionic composition, etc. can be varied in unique combinations to influence pathway performance. While CFME (i.e., selective enriching or functionalizing the lysate with pathway enzymes prior to extract generation) provides us with a rather quick way to screen enzyme ensembles and reaction conditions, this approach is limited by the cell's ability to produce the enzymes individually *in vivo*, a limitation that we address below.

3.4.2 Cell-Free Protein Synthesis Driven Metabolic Engineering

We next aimed to combine CFPS and CFME to modularly build the *n*-butanol pathway for forward engineering. This is dissimilar from previous works in which synthetic *in vitro* pathways have been built by purified enzymes or using lysates selectively enriched by heterologously expressed enzymes. Integration of CFPS enables one to speed up DBT cycle time for prospecting biosynthetic pathways. Indeed, using CFPS to express enzymes can reduce the time to build pathways to hours rather than days. As a proof-of-concept of this approach, we tested making Hbd2 (the non-native entry enzyme to the pathway) by CFPS to activate *n*-butanol production (**Figure 3.8A**). The key idea of the experiment was that the pathway would remain inactive (as downstream enzymes will not have their substrates) until active Hbd2 was synthesized. We chose to validate CFPS-ME in a three-step process. First, we quantified our ability to express Hbd2 in a CFPS reaction comprised of a mixture of lysates harboring selectively enriched pathway enzymes lacking Hbd2. This was important because typical CFPS systems use lysates from cells harvested in mid-late exponential phase, where as our lysates were collected 4 h post-induction of pathway enzymes. Second, we studied the ability to activate the entire pathway by combining CFPS and CFME. Third, we carried out a series of optimizations to try to increase yields.
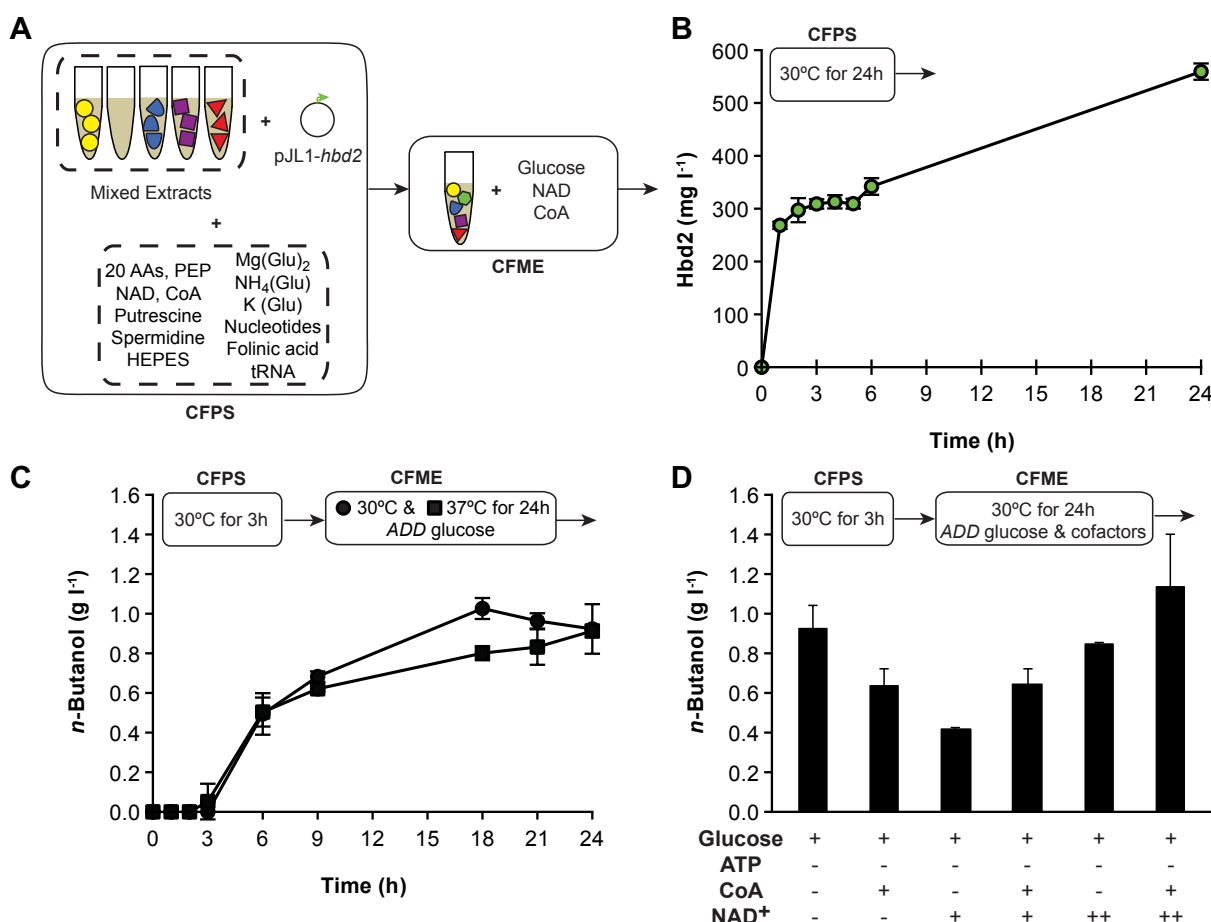
Figure 3.8. Cell-free protein synthesis of entry enzyme activates n-butanol production in vitro by CFPS-ME approach. *(A) Diagram describing the CFPS-ME experimental design. (B) Cell-free protein synthesis titers of Hbd2 from pJL1-hbd2 in a crude lysate mixture containing AtoB (EC), Crt1 (CA), Ter1 (TD), and AdhE1 (CA) overexpressed as determined by radioactive ^{14}C-leucine incorporation. CFPS reactions incubated over a 24-hr period at 30 °C. (C) n-butanol production in the same mixed lysate system activated by cell-free protein synthesis of Hbd2 run at 30 °C for 3 h. Glucose was added to activate the n-butanol pathway and CFME reactions were incubated over a 24-hr period at both 30 °C and 37 °C. (D) Cofactor (ATP, CoA, NAD+) optimization of downstream (ME portion of the CFPS-ME approach) cell-free reactions producing n-butanol were performed. Minus (-) signs represent no cofactor added, plus (+) signs represent mM amounts of cofactor to match conditions in CFME-alone experiments, and plus-plus (++) reactions represent double the amount of that cofactor. Reactions incubated for 24 h at 30 °C. All error bars represent standard deviations with n ≥ 3 independent reactions.*

For CFPS, we used the tunable and well characterized PANOx-SP CFPS system developed by Jewett and Swartz[152] to quantitatively test the synthesis of Hbd2. CFPS reactions at 30 °C were allowed to run for 24 h in batch operation and the yields of cell-free synthesized Hbd2 was quantified by monitoring ^{14}C-leucine incorporation. We based the system on a mixture of lysates used above, except the lysate with Hbd2 was not included. Endogenous protein synthesis machinery should act to synthesize and fold desired

protein products upon incubation with essential substrates (*e.g.,* amino acids, nucleotides, DNA or mRNA, energy substrates, cofactors, and salts). In this case, we showed that when the DNA for the Hbd2 enzyme on a pJL1 vector was added, the mixed extract could produce 559 ± 15 mg l^{-1} of Hbd2 over a 24-hour period (**Figure 3.8B**). Based on this result and the fact that this reaction was over 50% complete by three hours, we chose to run all subsequent CFPS reactions for three hours, which should provide sufficient protein quantities for prototyping.

We next investigated the ability of the cell-free synthesized Hbd2 to activate the full *n*-butanol pathway. After three hours of CFPS, we initiated *n*-butanol metabolism by adding 200 mM glucose to the reactions. We showed that CFPS of Hbd2 could activate *n*-butanol metabolism reaching a titer of 0.92 ± 0.13 g l^{-1} (**Figure 3.8C**). Negative control reactions without synthesis of the Hbd2 did not produce *n*-butanol. Notably, the CFME portion resulted in the same *n*-butanol yields when carried out at either 30 or 37 ºC, so for ease we selected 30 ºC for all future experiments to have the CFPS and CFME portions performed at the same temperature. As in the CFME system alone, we found that small molecules, cofactors, etc. can modulate pathway performance. For example, we found that adding both NAD and CoA with glucose to initiate *n*-butanol metabolism after CFPS gave us 1.22 ± 0.22 g l^{-1} *n*-butanol (**Figure 3.8D**). Collectively, our results prove for the first time to our knowledge the ability to combine CFPS and CFME to support a highly active biosynthetic pathway.

We further extended this proof-of-concept to activate *n*-butanol production using CFPS at any pathway node by producing each *n*-butanol pathway enzyme. Using mixed extracts with all but one necessary enzyme, we performed CFPS of the 'missing enzyme' and saw that each enzyme could be produced individually at more than 100 mg l^{-1} without optimization (**Figure 3.9A**). We then proved that full product of each protein is made exclusively in each reaction by an autoradiogram (**Figure 3.10**). After validating expression of each enzyme, we then performed CFPS-ME reactions. We carried out three-hour CFPS reactions and then initiated the *n*-butanol pathway by adding glucose, NAD, and CoA, because supplementation of CFPS-ME reactions with both NAD and CoA increased *n*-butanol titers for Hbd2 (**Figure 3.8D**). Strikingly, CFPS-ME could be used for each of the pathway enzymes to produce *n*-butanol at levels as high as 1.71 ± 0.06 g l^{-1} (**Figure 3.9B**).
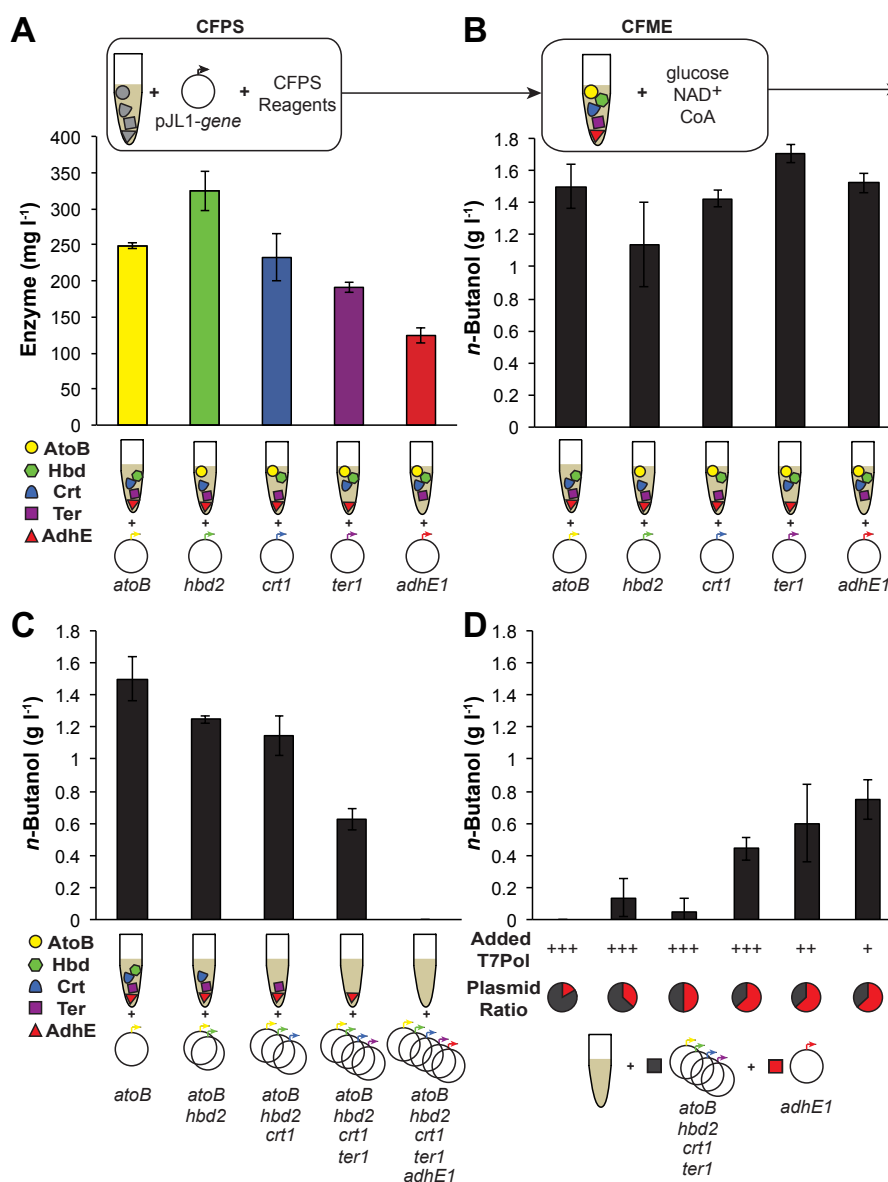
Figure 3.9. Using cell-free protein synthesis to activate metabolism from any node in the biosynthetic pathway. *(a) Cell-free protein synthesis titers of AtoB (EC), Hbd2 (CB), Crt1 (CA), Ter1 (CA), and AdhE1 (CA) off pJL1 constructs in separate reaction mixtures as determined by radioactive ¹⁴C-leucine incorporation. Each reaction mixture contained crude lysates with all pathway enzymes except the one made by CFPS. CFPS reactions were incubated for 3 h at 30 ºC. (b) n-butanol production in the same mixed lysate system activated by CFPS of each enzyme run at 30 ºC for 3 h. Glucose, CoA, and NAD⁺ were added to activate the n-butanol pathway and CFME reactions were incubated for 24 h at 30 ºC. (c) n-butanol production activated by CFPS of enzymes in combinations: (1) AtoB (EC); (2) AtoB (EC) and Hbd2 (CB); (3) AtoB (EC), Hbd2 (CB), and Crt1 (CA); (4) AtoB (EC), Hbd2 (CB), Crt1 (CA), and Ter1 (TD); and (5) AtoB (EC), Hbd2 (CB), Crt1 (CA), Ter1 (TD), and AdhE1 (CA). The CFPS reactions were run at 30ºC for 3 hrs. Glucose, CoA, and NAD⁺ were added to activate the n-butanol pathway and reactions were incubated for 24 h at 30 ºC. (d) A plasmid ratio optimization of pJL1-adhE1 vs. all other pJL1 constructs along with a test of three concentrations of T7 polymerase. For each, CFPS was run at 30 ºC for 3 h. Glucose, CoA, and NAD⁺ were added to activate the n-butanol pathway and reactions were incubated for 24 h at 30 ºC. All error bars represent standard deviations with n ≥ 3 independent reactions.*

Figure 3.10. SDS-PAGE and autoradiogram for CFPS of each individual enzyme show full-length protein formation. *(A) SDS-PAGE was run with two replicates of each CFPS reaction performed at standard conditions listed in the methods section and for 3 h at 30 ºC producing ^{14}C-Leu incorporated protein corresponding to each enzyme in the* n-*butanol pathway. Molecular weights are listed at the bottom of the lanes. (B) An autoradiogram of the same gel showing that each enzyme is expressed* in vitro *as full-length proteins.*

We next set out to demonstrate we could build the entire pathway by CFPS of the pathway enzymes in our extracts. To this end, we extended the number of enzymes made *in vitro* one by one, by adding equal amounts of DNA of each, and saw that when we made one, two, three, and four of the five enzymes necessary *in vitro* we could produce *n*-butanol at levels between ~0.6 and ~1.4 g l^{-1} (**Figure 3.9C**). Again, full-length product of each protein is made in each reaction as shown by autoradiogram (**Figure 3.11**). However, as we increase the number of enzymes produced by CFPS, the amount of *n*-butanol synthesized decreases. In fact, when we tried to produce all five enzymes *in vitro* we were initially unable to make any *n*-butanol. We attribute this drop-in *n*-butanol production to there not being enough of the last enzyme in the pathway, AdhE, seen by quantification of the enzymes produced by CFPS (**Figure 3.12).** However, we were able to make all enzymes *in vitro* at sufficient levels necessary to make *n*-butanol at 0.75 ± 0.12 g l^{-1} by increasing the plasmid DNA encoding AdhE to more than 50% of the total DNA added, (**Figure 3.9D; Figure 3.13**). Reduced T7 polymerase added shows improvements in *n*-butanol production. Typical CFPS systems supplement T7 polymerase stored in glycerol and increasing glycerol concentrations can be

deleterious to the CFPS system. The extract used in this study contains T7 polymerase expressed *in vivo* prior to extract preparation, so T7 polymerase in the extract is expected to be sufficient without supplementation. Based on our result that added ATP was deleterious to *n*-butanol production by CFME (**Figure 3.4B**), the ATP used in CFPS might be expected to inhibit CFPS-ME *n*-butanol titers if ATP is long-lived. We have previously shown that ATP concentrations are stable around 200 µM over a ~6-8 hour batch CFPS reaction[153]. Though, a negative effect from ATP is expected, it is difficult to use the CFME optimization conditions for CFPS-ME, given the added complexity of protein synthesis. Our results importantly showed that we could build a five-step heterologous pathway to make *n*-butanol *in vitro* in three hours.

Figure 3.11. Protein SDS-PAGE gel and autoradiogram for CFPS of multiple enzymes produced full length product of each protein. *SDS-PAGE was run with CFPS reactions containing one, two, three, four, and five DNA plasmids, building up the pathway. Reactions were each run at standard conditions listed in the methods section and for 3 h at 30 ºC producing ^{14}C-Leu incorporated protein corresponding to each enzyme in the* n-*butanol pathway.*

Figure 3.12. Quantification of individual enzymes in multiple protein CFPS reactions from autoradiogram indicates a lower yield of downstream pathway enzymes. *(A) BL21(DE3) extract containing overexpressed proteins indicated in the key, different DNA plasmid compositions, and standard CFPS reagents identified in the methods section were incubated at 30 ºC for 3 h. Each sample was separated by SDS-PAGE and stained with Coomassie blue. The gel was exposed to autoradiogram showing full-length product of each enzyme resulting from the DNA combinations. (B) Using densitometry with ImageJ software, each lane from the gel in panel A was analyzed for band density to determine approximate, relative amounts of each pathway enzyme produced by CFPS. The bars represent the percent of each enzyme produced* in vitro. *After CFPS for 3 h at 30 ºC, each reaction was supplemented with glucose, NAD+, and CoA and incubated at 30 ºC for an additional 24 h to measure* n-butanol production *from each mix. Orange circles indicate* n-butanol titers with error bars representing 1 s.d. with n ≥ 3.*

Figure 3.13. Plasmid optimization for CFPS-ME of all pathway enzymes in vitro shows the ability to produce n-butanol. *CFPS reactions were run in BL21(DE3) extract containing no overexpressed proteins. DNA plasmids encoding each heterologous enzyme were added in equal ratios with pJL1-adhE1 modulated as per the divisions in the circular graphics in the figure key. In the key, black represents equal amounts of plasmids encoding AtoB, Hbd, Crt, and Ter, and red represents the amount of pJL1-adhE1. CFPS reagents were added and incubated at 30 ºC for 3 h. Glucose, NAD+, and CoA were added, and samples were further incubated at 30 ºC for 24 h.* n-Butanol production is measured and differs with varied concentrations *of added T7 polymerase. All error bars represent 1 s.d. with n ≥ 3.*

3.4.3　Rapid Prototyping and Enzyme Discovery with CFPS-ME

The ability to use CFPS-ME to produce enzymes for *n*-butanol biosynthesis allows us to test pathway enzymes without expressing enzymes in the host cell. As a model case study, we decided to test for improved pathway performance (increased *n*-butanol production) by swapping out some of our initial Ter and AdhE enzymes for a variety of homologs. In less than a day, we studied 4 Ter and 3 AdhE homologs in a combined CFPS-ME reaction. In all cases, we observed synthesis of *n*-butanol, though lower than our previous best-performing enzymes (**Figure 3.14A**). Five of these variants come from species never tested before.

Figure 3.14. Using CFPS-ME to rapidly screen pathway enzymes*. (A)* n-*butanol production activated by CFPS of unique Ter homologs and AdhE homologs from pJL1 constructs: Ter3 (*Fibrobacter succinogenes, FS*), Ter4 (*Flavobacterium johnsoniae, FJ*), Ter5 (*Spirochaeta bajacaliforniensis, SB*), Ter6 (*Cytophaga hutchinsonii, CH*), AdhE9 (*Thermosynechococcus sp. NK55a, TN*), AdhE10 (*Providencia burhodogranariea, PB*), and AdhE13 (*Serratia marcescens, SM*). Ter homologs were expressed in crude lysate mixtures containing AtoB (EC), Hbd2 (CB), Crt1 (CA), and AdhE1 (CA) overexpressed, and AdhE homologs were expressed in lysates containing AtoB (EC), Hbd2 (CB), Crt1 (CA), and Ter1 (TD) overexpressed. (B) n-Butanol production activated by CFPS putative bifunctional enzymes for Hbd and Crt activity: Hbdcrt2 (*Aeropyrum camini, AC*), Hbdcrt3 (*Pyrobaculum aerophilum, PA*), Hbdcrt4 (*Sulfolobus islandicus, SI*), and Hbdcrt6 (*Sulfolobus acidocaldarius, SA*). CFPS reactions were performed from linear DNA in crude lysate mixtures containing: (1) AtoB (EC), Ter1 (TD), and AdhE1 (CA) overexpressed to test bifunctionality, (2), AtoB (EC), Crt1 (CA), Ter1 (TD), and AdhE1 (CA) overexpressed to test Hbd functionality alone, and (3) AtoB (EC), Hbd2 (CB), Ter1 (TD), and AdhE1 (CA) overexpressed to test Crt functionality alone. For each, CFPS was run at 30 ºC for 3 h. Glucose, CoA, and NAD+ were added to activate the* n-*butanol pathway and reactions were incubated for 24 h at 30 ºC. All error bars represent standard deviations with n ≥ 3 independent reactions.*

Having demonstrated the ability to explore enzyme homologs using CFPS-ME, we then set out to demonstrate the potential for using linear DNA templates instead of plasmids. Using linear DNA molecules, i.e. PCR products, would expedite the process since the entire process could be done without cells and we could avoid laborious cloning steps. As a model system, we first repeated the experiments presented in **Figure 3.14** with linear templates and observed that the linear DNA templates can successfully be expressed to complete the *n*-butanol biosynthesis pathway (**Figure 3.15**). Next, we chose to screen multifunctional enzymes that to our knowledge have never before used for *n*-butanol production. We

selected four enzymes with proposed Hbd and Crt functionalities that were identified by NCBI-BLAST searches. By preparing reactions with three different enzyme mixtures (mixed extracts with overexpressed enzymes prior to lysis) (1) without Hbd, (2) without Crt, and (3) without Hbd and Crt, we could characterize each enzyme variant by their ability to perform each enzymatic function. We discovered that each of these enzymes could activate *n*-butanol synthesis, and the proposed Hbdcrt6 from *Sulfolobus acidocaldarius* only had Hbd functionality (**Figure 3.14B**). The ability to use linear DNA templates for CFPS-ME makes possible the ability to rapidly screen individual and sets of enzymes completely *in vitro*. Here, we used this approach to parse out individual functionalities of multi-functional enzymes.
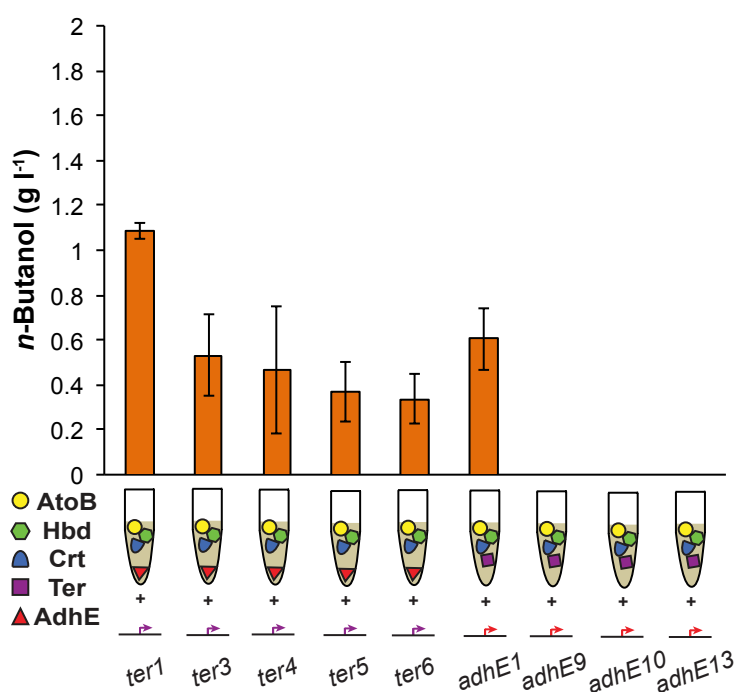
Figure 3.15. Using CFPS-ME to rapidly screen pathway enzymes from linear DNA templates. *Linear DNA of each enzyme variants were created with the regulatory elements from pJL1 and with a randomized ~20 bp on each end and used as DNA template in each CFPS reaction. CFPS of each enzyme variant (Ter and AdhE) was used to activate* n-*butanol production in crude lysate mixtures containing AtoB (EC), Hbd2 (CB), Crt1 (CA), and AdhE1 (CA) overexpressed or AtoB (EC), Hbd2 (CB), Crt1 (CA), and Ter1 (TD) overexpressed, respectively. All error bars represent 1 s.d. with n ≥ 3.*

3.5 Discussion

In this study, we developed a new cell-free framework for prototyping biosynthetic pathways and screening enzymes. In one scenario, we overexpress individual pathway components in cells, lyse these cells, and mix and match lysates in cell-free cocktails to study biochemical pathway performance. In a distinct thrust from typical *in vitro* systems, our approach allows us to study heterologous pathways in the context of native metabolism. In another scenario, we bypass *in vivo* expression altogether by using CFPS to enrich lysates with different enzymes for combinatorial assembly of different pathways. The combination of CFPS to express homologs of individual biosynthetic enzymes for studying pathway performance is also a distinction of our workflow. In addition, the use of linear PCR templates, which could be improved by DNA stability techniques (*e.g.,* the addition of purified GamS protein)[65], allows us to avoid *in vivo* cloning steps altogether. Our CFPS-ME approach should therefore be faster than conventional approaches to select enzymes and pathway designs in cells (hours instead of days/weeks) and enables parallelized pathway construction of combinatorial designs to accelerate DBT cycles.

A key conceptual innovation of our work is that the DBT unit can be cell-free lysates rather than genetic constructs. Engineering large biosynthetic systems composed of many genes in microbes remains challenging[142]. One of the many obstacles is simply how many different genetic designs with beneficial chances are feasible to make. Cell-free systems have already been shown to screen genetic designs to improve enzyme performance at a rapid rate[155]. Our CFPS-ME framework should allow researchers to study more designs than previously possible by rapidly prototyping enzyme performance *in vitro* before putting designs into a host. As an example, a six-step biosynthetic pathway testing 5 homologs for each enzymatic step would require testing of 15,625 pathway combinations. While this set of combinations exceeds typical pipelines pursued in cells today, our CFPS-ME system could leverage robotic or automated liquid-handling systems to access such design space.

The goal of this manuscript was to provide a new approach to building biosynthetic pathways in a modular fashion *in vitro.* Now achieved, we plan to optimize a large-scale fermentation process with the CFPS-ME approach in the future. Towards this goal, we additionally carried out experiments to show that protein expression in the cell-free system translates to the *in vivo* system. Specifically, we took all Ter

homologs screened *in vitro* by CFPS-ME (**Figure 3.14**) and expressed them in whole cells *in vivo*. All but

one of the Ter homolog proteins can be expressed in cells on a first pass (as determined by SDS-page

expression, **Figure 3.16**). These data show that protein expression in the cell-free system can translate to

the *in vivo* system. Thus, our approach holds promise for identifying good enzymes that can be expressed

in cells, following a body of work that uses *in vitro* enzyme assays to identify enzymes with the best-

performing biochemical characteristics for desired metabolic transformations prior to putting them into a

host. For example, Liao and colleagues showed that *in vitro* reconstitution could be used to construct the

non-oxidative glycolytic pathway prior to *in vivo* expression[46], and Zhu *et* al. reconstituted the mevalonate

pathway *in vitro* to study pathway kinetics before using the pathway *in vivo* for the production of farnesene[44].
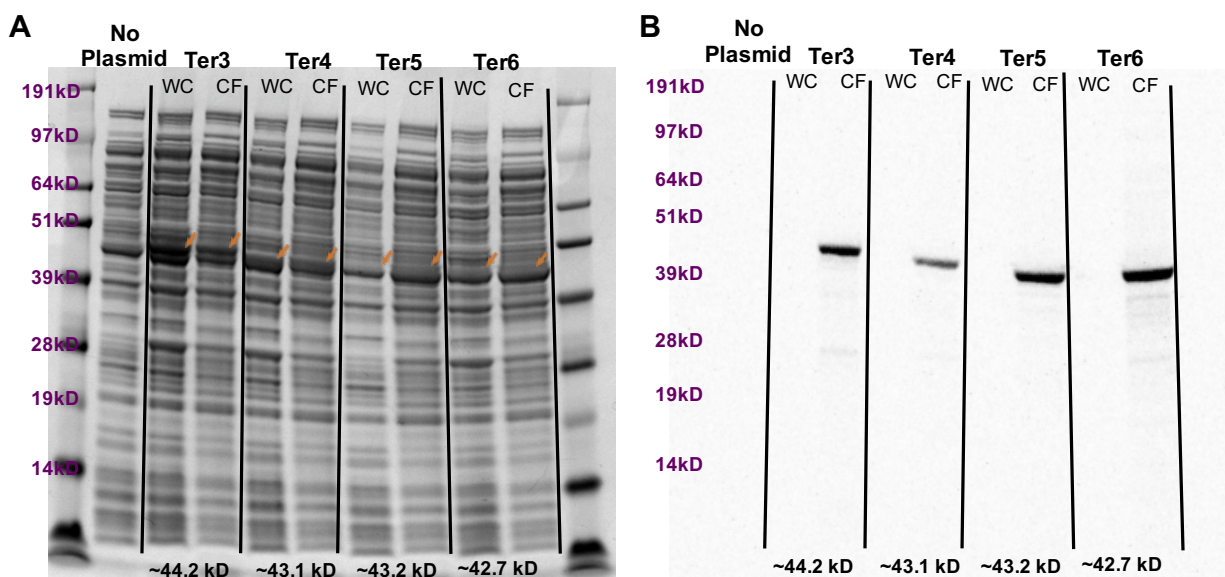
Figure 3.16. Expression of Ter homologs in cells versus cell-free. *SDS-PAGE was run on samples expressing the Ter homologs screened in Figure 6 of the manuscript. The first lane is the SeeBlue Plus2 Pre-stained standard. The second lane is a no plasmid (negative) control in BL21(DE3). For each homolog, the WC lane is a whole cell sample taken 4 hours after induction, and the CF lane is a cell-free protein synthesis sample taken after 3 hours of expression. (A) Coomassie-stained gel. While expression of Ter5 in the WC lane in an un-optimized (generic RBS) single attempt is unobservable, the other Ter enzymes are expressed, which confirms that our approach holds promise for identifying good enzymes that can be expressed in cells. (B) An autoradiogram of the same gel showing that each enzyme is expressed in vitro. As the* in vivo *expression did not incorporate radioactive leucine, bands are not expected in the WC lanes in panel B.*

Our cell-free approach mimics the intracellular environment of *E. coli,* where endogenous glycolytic

enzymes from the cell extract convert glucose to AcCoA. Thus, our platform enables many different

biosynthetic pathways to be studied in the context of central metabolism with enhanced control inherent to *in vitro* systems. Here, we were able to increase *n*-butanol production by ~200% of our initial starting conditions (up to ~1.5 g l[-1]) by simply testing the performance of different enzymes sets and adjusting the physicochemical environment. While it is be difficult to compare in a normalized fashion the *in vitro* process to the *in vivo* process, our results (given as final measured concentration) are higher than some published reports of *n*-butanol production in comparable genomically unmodified hosts (**Figure 3.17**)[57,156-158]. However, Bond-Watts *et* al. notably reported titers of 4.6 g l[-1] in a genomically unmodified host by selecting a particular set of synergistic enzymes and taking advantage of their chemistries[137]. Given the reasonable yields, we were curious as to how the CFPS-ME reactions would perform at increased scale. We thus performed additional experiments of increasing size reactions to give confidence in our quantitative yields. Specifically, the reaction volume of CFPS-ME reactions was scaled from 25 to 250 µL, an order of magnitude increase **(Figure 3.18).** Our data shows that these reactions are scalable and are consistent with several previous works showing the ability of cell-free systems to scale linearly[121,159-162].
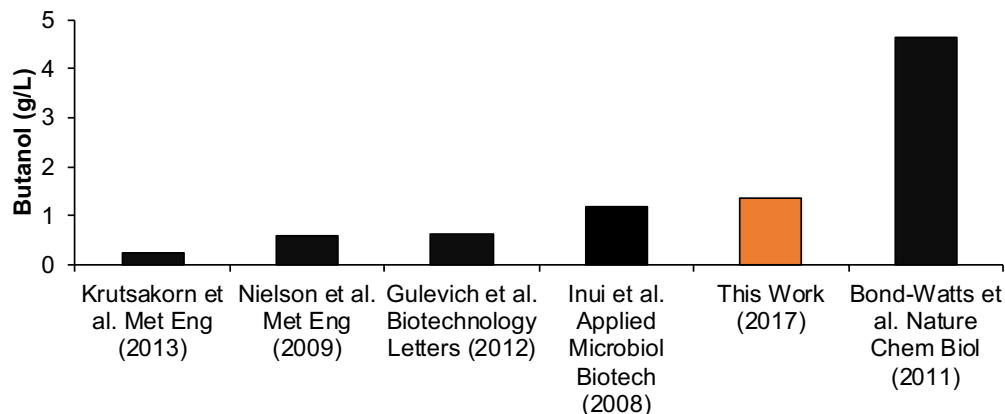
Figure 3.17. Comparison of n-butanol production in genomically unmodified hosts. *Final titers (g/L) of* n-*butanol are reported from recent studies using genomically unmodified host strains. The Krustakorn* et al. *2013 study is the only other* in vitro n-*butanol production system (uses purified enzymes). The highest titer from this study is reported in orange.*
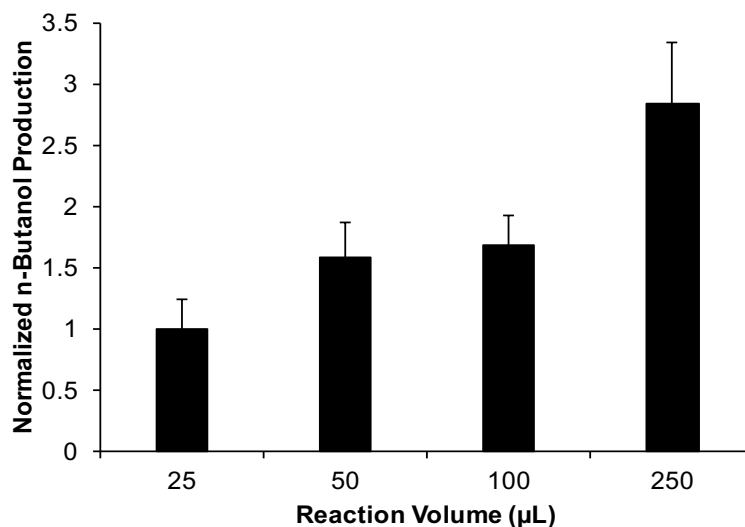
Figure 3.18. Scale-up of CFPS-ME reactions. *CFPS reactions were run in BL21(DE3) extract containing no overexpressed proteins in 1.5mL Eppendorf tubes. DNA plasmids encoding each heterologous enzyme were added in equal ratios with pJL1-adhE1 representing 70% of the total DNA. CFPS reagents were added and incubated at 30 ºC for 3 h. Glucose, NAD+, and CoA were added, and samples were further incubated at 30 ºC for 24 h. The total reaction volumes are listed on the x-axis with the data normalized to 25µL reactions. Increasing the reaction volume improved overall butanol yields. All error bars represent 1 s.d. with n ≥ 3.*

Looking forward, specialty chemicals, natural products, and materials offer an extremely diverse set of compounds with a seemingly infinite set of structures and bioactivities. Our CFPS-ME approach offers a new discovery pipeline to leverage advances in DNA sequencing and DNA synthesis to optimize biosynthetic pathways, discover new enzymes, and test new hypotheses. Because it is an open system, cell-free reactors can be readily interrogated for intermediate product formation, such as by the on-line, high speed LC/MS approaches used by Panke and colleagues for optimization of glycolysis in cell-free extracts[72]. Cell-free systems in tandem with high-end metabolomics could offer a high degree of flexibility to model the kinetics and stability of individual enzymes, measure metabolite fluxes in multistep pathways, and experimentally isolate many other parameters confounded in living organisms. This has potential to speed up metabolic engineering DBT cycles.

3.6 Conclusions

This work is the first development of cell-free protein synthesis for enzymatic pathway construction and has spawned a growing interest in cell-free metabolic systems. This chapter documented how this

framework was designed, how it can be used to assemble an entire biosynthetic pathway, and how it can be used to discover enzyme functionalities. This chapter also played a major role in my development as a confident scientist. This work also hints at being able to speed up DBT cycles for metabolic engineering, a feat that will be addressed in Chapter 7.

Chapter 4: Methods for *in vitro* prototyping and rapid optimization of biosynthetic enzymes

---

Try not to become a man of success, but rather try to become a man of value.

- Einstein

This chapter is one of my last pieces of work during my PhD for two reasons: (1) the invitation to write was received based on the interest garnered from our previous work in pathway prototyping (i.e. Chapter 3) and (2) it was an important culmination of the research I performed. If I was going to profess the creation of cell-free framework that would actually shift the community's way of thinking around designing, building, and testing biosynthetic pathways, I needed to supply details as to how everyone can build pathways in cell-free systems. Thus, Chapter 4 was written and published in a volume of Methods in Enzymology.

4.1 Abstract

Engineering biological systems for the production of biofuels and bioproducts holds great value but often requires laborious, time-consuming design-build-test cycles. For decades cell-free systems have offered quick and facile approaches to study enzymes with hopes of informing cellular processes, mainly in the form of purified single-enzyme activity assays. Over the past twenty years, cell-free systems have grown to include multi-enzymatic systems, both purified and crude. By de-coupling cellular growth objectives from enzyme pathway engineering objectives, cell-free systems provide a controllable environment to direct substrates toward a single, desired product. Cell-free approaches are being

developed for prototyping and for biomanufacturing. In prototyping applications, the idea is to use cell-free systems to select and optimize biosynthetic pathways for informing cellular design. We present a detailed method for the generation of S12 crude lysates for cell-free pathway prototyping, mix-and-match cell-free metabolic engineering using pre-enriched lysates, and cell-free protein synthesis driven cell-free metabolic engineering. The cell-free synthetic biology methods described herein are generalizable to any biosynthetic pathway of interest and provide a powerful approach to building pathways in crude lysates for the purpose of prototyping. The foundational principle of the presented approach is that we can construct discrete metabolic pathways through modular assembly of cell-free lysates containing enzyme components produced by overexpression in the lysate chassis strain or by cell-free protein synthesis (CFPS; *in vitro* production). Overall, the modular and cell-free nature of our pathway prototyping framework is poised to facilitate multiplexed, automated study of biosynthetic pathways to inform systems-level design.

4.2   Introduction

Increased demands for energy, climate change concerns, and reliance on petrochemicals as the source of 95% of today's chemicals and materials have intensified the need for sustainable, low-cost biofuels and bioproducts production [1,2]. Microbial cell factories offer one of the most attractive approaches for addressing this need [7]. However, long research and development timelines (10s-100s of person years) motivates the need for new methods to accelerate the design and optimization of biological systems [4,7,124-127].

4.2.1   The state of metabolic engineering

Often, biologically-produced small molecules are insufficient for production at commercially relevant titers, rates, or yields because natural sources are difficult to optimize and to scale. Thus, engineers seek to design enzymatic reaction schemes in model microorganisms to meet manufacturing criteria. Success in these endeavors depends upon identifying sets of enzymes that can convert readily available molecules (e.g., glucose) to high-value products (e.g., medicines), with each enzyme performing one of a series of chemical modifications. For example, introducing heterologous pathways into model microbes

(such as baker's yeast and *Escherichia coli*) and engineering them to maximize biochemical production has led to large-scale production of 1,3-propanediol, farnesene, and artemisinin with many more on their way to market [4,17]. Unfortunately, this is difficult because design-build-test (DBT) cycles—iterations of re-engineering organisms to test new sets of enzymes—can be detrimentally slow due to the constraints of cell growth [163]. As a result, a typical project today might only explore dozens of variants of an enzymatic reaction pathway. This is often insufficient to identify a commercially relevant solution because selecting productive enzymes using existing single-enzyme kinetic data has limited applicability in multi-enzyme pathways and consequently requires more DBT iterations. While techniques continue to develop to multiplex DBT cycles for rationally engineering cells [142], *in vitro* systems show promise in speeding up DBT cycles because they bypass many *in vivo* limitations by having direct access to the cellular contents [17,64,65,71]. In concert with a trend toward automation, *in vitro* systems could transform the way we engineer metabolic pathways.

4.2.2   Emerging cell-free biotechnology

Cell-free systems complement traditional cellular systems. By de-coupling cellular growth objectives from engineering enzyme utilization objectives, cell-free systems provide controllable and direct environment to direct substrates toward a single, desired product [144]. Thus, cell-free synthetic biology methods are being developed for both pathway prototyping and for biomanufacturing. In prototyping applications, the idea is to use cell-free systems to select and optimize biosynthetic pathways for informing cellular design. Already, cell-free biosynthetic pathway building methods are already being used for pathway operation and debugging [144]. However, using cell-free systems to study metabolism and enzymatic pathways are still in its early stages. Nevertheless, cell-free systems provide advantages in controlling enzymes and the reaction environment. In biomanufacturing applications, the idea is to use cell-free systems for making the product itself. There is also potential to make products unavailable to cells due to toxicity limitations and to focus substrates to products with yields unattainable in cells [60,164]. Cell-free systems might also offer exciting new directions in the synthesis of hybrid biochemicals comprised of parts derived from organic syntheses and parts derived from biological syntheses [19]. The ability to not only use enzymes from multiple organisms but also unique metabolisms from across the phylogenic spectra are also

on the horizon. As CFME emerges [23,144,146,165], two broad classes dominate *in vitro* small molecule synthesis: purified enzymes and crude cell lysates.
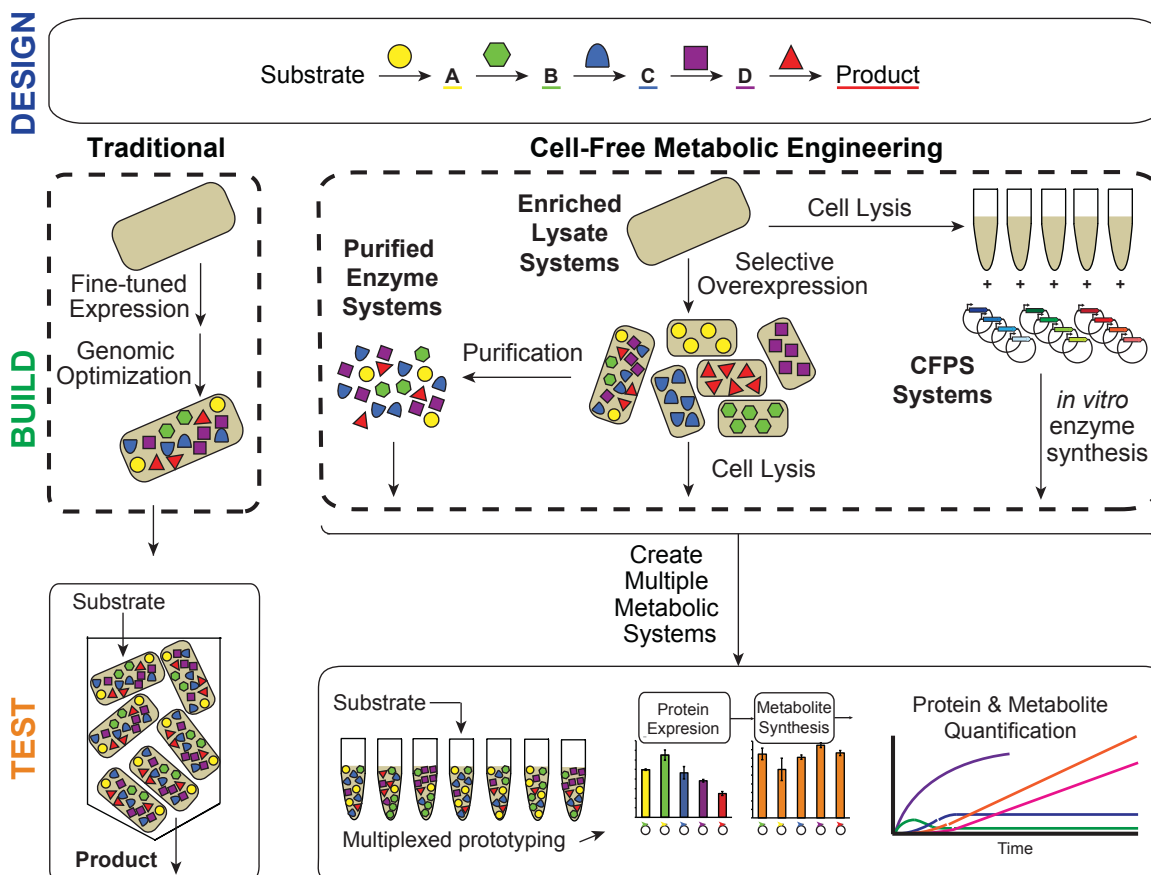
Figure 4.1. Overview of cell-free synthetic biology methods for prototyping biosynthetic pathways.

4.2.2.1 Purified enzyme systems

Purified enzyme approaches involve individual overexpression and purification of enzymes, which are then used as individual biocatalysts or recombined to assemble pathways of interest. The benefit of these systems is that the reaction network is explicitly defined, which gives exquisite control of reaction conditions and pathway fluxes. Indeed, the clarity of the biosynthetic pathway comes from eliminating unnecessary enzymes and cellular interferences (*i.e.,* growth, other off-pathway metabolites). There are several examples of simple purified enzyme systems in industrial biocatalysis. [48,166]. However, few industrial examples of synthetic enzymatic pathways exist, in part because of the high catalyst and cofactor costs (as

a result of enzyme purification and instability) and poor cofactor regeneration [49]. Despite these challenges, the majority of CFME research to date has utilized purified systems [44,53-56,84,167]. Recently, unique cofactor regeneration systems have been developed for purified systems that increase production capabilities and longevity of reactions [103,168].

4.2.2.2   Crude cell lysate systems

Crude extract-based systems rely on the ensemble of biocatalysts left after cell lysis. Upon centrifugation lipid membranes and genomic DNA are removed from cell lysates. Many groups have carried out extensive work in extract preparation and system optimization [51,152,153,169]. There are several advantages to using crude lysates beyond their ease of preparation including lower system catalyst costs compared to purified counterparts, cofactor regeneration systems [19,37], and the presence of native-like metabolism [51,152]. Crude lysates thus allow for observations of metabolic interactions with biosynthetic pathways [121,170,171]. There are a growing number of successes in using and characterizing crude lysate systems. A great example is the real-time monitoring and optimization of DHAP production [72,172]. In addition, our group has shown that 2,3-butanediol [121], mevalonate [171], *n*-butanol [170,173], and more complex products [174] can be constructed in crude lysates with high productivities (>g/L/h).

4.2.3   The cell-free metabolic engineering framework

The cell-free framework is a way to build pathways in the context of the Design-Build-Test paradigm. The foundational principle of the cell-free metabolic engineering approach is that we can construct discrete metabolic pathways through modular assembly of cell-free lysates containing enzyme components produced by overexpression in the lysate chassis strain or by cell-free protein synthesis (CFPS; *in vitro* production) (**Figure 4.2**). In addition, the open reaction environment allows for the supplementation of components such as cofactors and intermediates at any time during a cell-free reaction, which can be maintained at precise concentrations. In the Design phase, desired small molecule products are selected, enzymes, homologs, and their stoichiometries are chosen, and likely beneficial conditions (i.e., substrates, cofactors, buffers, pH, and temperature) are selected. The Build phase, the heart of this framework, consists of assembling planned pathways for making the desired molecules from lysates, which can occur through two routes. One pathway construction route we term mix-and-match cell-free metabolic

engineering (CFME) involves preparing multiple cell extracts for a selected pathway. One enzyme in the candidate pathway will be pre-enriched in each extract by overexpression in the host strain prior to lysis. Then, enriched extracts can be mixed in multiple, different ratios to build complete biosynthetic pathways. Another route termed CFPS driven metabolic engineering (CFPS-ME) involves one-pot *in vitro* synthesis of biosynthetic enzymes and pathway operation. In this route, enzymes are made by CFPS in the extract post-lysis by adding the DNA for each enzyme along with protein synthesis reagents followed by the addition of pathway operation reagents, faster than any previous approach (hours rather than days). Pathways can be built in 96- or 384-well plates, using liquid handling robotics, or in 1.5- or 2.0-mL micro-centrifuge tubes [170]. Significantly, this cell-free prototyping approach does not require the focus on flux balancing and delicate promoter tuning to maintain viability as is true for *in vivo* systems [175-178]. Following the rapid and discrete construction of metabolic pathways, each pathway can be tested by using kinetically sampled batch reactions for fast DBT cycle times [173].
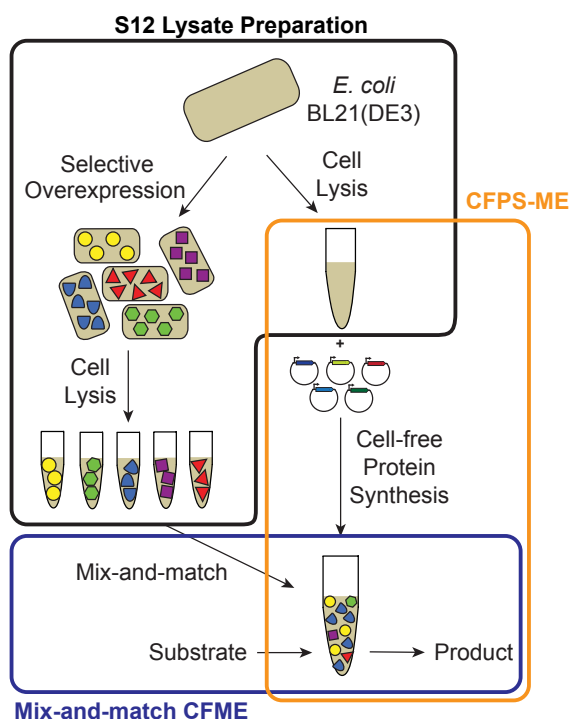
Figure 4.2. Overview of methods for cell-free pathway prototyping. *Lysate preparation, mix-and-match CFME and CFPS-ME methods are outlined.*

4.3 S12 lysate preparation for cell-free metabolic engineering

The foundation of these cell-free prototyping strategies relies on crude lysates for pathway assembly. There are many ways to make *E. coli*-based crude lysates [169,179-183]. In theory any of these methods would work to make lysates for pathway engineering, so our method focuses on one of the more high-throughput, high-yielding methods for crude lysate preparation [169]. The following protocol is based on Kwon & Jewett's preparation with a few modifications for pathway prototyping strategies, also described in Karim and Jewett's work [173].

4.3.1 Materials

4.3.1.1 Equipment

- 125 mL (small batch) and 4L (large batch) baffled, Tunair shake flasks, or similar culture container; one per extract.

- 30 and 37 °C incubators with vigorous shaking at 250 rpm.

- Spectrophotometer, for cuvette-based OD_{600} measurements.

- Optically clear cuvettes for measuring optical density.

- Centrifuge capable of spinning 1 L centrifuge bottles at 5,000xg, pre-chilled to 4 ºC.

- 1 L plastic centrifuge bottles and lids, pre-chilled to 4 ºC; one per extract.

- Table-top centrifuge capable of spinning 50 mL conical tubes at 12,000xg, pre-chilled to 4 ºC.

- 50 mL conical tubes, pre-chilled to 4 ºC; one per extract.

- Weight scale to measure wet cell mass.

- Kim wipes.

- Bucket with ice.

- Beaker with ice water bath.

- Liquid nitrogen and dewar.

- −80 °C Freezer for storage.

- Q125 Sonicator (Qsonica, Newtown, CT), 3.175 mm diameter probe at frequency of 20 kHz.

- Heat block capable of heating to 90 ºC.

- XCell SureLock® Mini-Cell (Invitrogen) with power box.

- Gel Doc™ XR+ Gel Documentation System (BioRad).

4.3.1.2   Media

- Luria–Bertani (LB) broth: 1.0% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, autoclaved.

- LB agar: add 1.5% (w/v) agar to LB prepared as above. Aliquot 30 mL of LB agar per 150 mm petri dish into a plastic conical tube. Add appropriate antibiotics.

- 2xYTPG broth: 1.6% (w/v) tryptone, 1.0% (w/v) yeast extract, 0.5% (w/v) NaCl, 1.8% (w/v) glucose, 0.7% (w/v) K_2HPO_4, 0.3% (w/v) KH_2PO_4, pH adjusted to 7.2 with 5 N KOH, autoclaved. Add 40% (w/v) glucose, separately autoclaved, to broth prior use. Use 1 L per extract.

4.3.1.3   Media Supplements

- Carbenicillin (Carb) (100 µg/ml): To make a 1,000x stock, mix 1 g in 10 mL nanopure water, sterile-filtered.

- Kanamycin (Kan) (50 µg/ml): To make a 1,000x stock, mix 0.5 g in 10 mL nanopure water, sterile-filtered.

- Isopropyl-β-D-thiogalactopyranoside (IPTG) (0.1 mM): To make a 1,000x stock, mix 0.238 g in nanopure water to 10 mL, sterile-filtered.

4.3.1.4   Bacterial Strains & Plasmids (Table 4.1)

- *E. coli* DH5α (NEB).

- *E. coli* BL21(DE3) (NEB).

- pETBCS-rbsU vector, used in previous studies.

- pJL1 vector, used in previous studies.

Table 4.1. A list of strains and plasmids used to do cell-free pathway prototyping.

Strain or Plasmid	Description	References
E. coli strains		
NEB Turbo™	F' proA+B+ lacIq ΔlacZM15 / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)TetS endA1 thi-1 Δ(hsdS-mcrB)5 -- used for cloning purposes only	New England Biolabs
BL21 (DE3)	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5 -- used for enzyme overexpression and for lysate production	New England Biolabs
Plasmids		
pETBCS-rbsU-gene	Plasmid used for protein production and extract preparation	[173]
pJL1-gene	Plasmid used for CFPS	[173]; Addgene #69496

4.3.1.5   Buffers and reagents

- S30 buffer: 10 mM Tris-acetate (pH 8.2), 14 mM magnesium acetate, and 60 mM potassium glutamate, pre-chilled to 4 ºC.

- QuickStart™ Bradford Protein Assay Kit (BioRad Laboratories, Inc.).

- NuPAGE® LDS Sample Buffer (Invitrogen).

- 1M Dithiothreitol (DTT).

- 4-12% Bis-Tris Nu-PAGE gel (Invitrogen).

- 20x NuPAGE MOPS SDS Running Buffer (Invitrogen).

- SeeBlue™ Plus2 Pre-stained Protein Standard (Thermo).

- SimplyBlue™ SafeStain (Thermo)

4.3.2   Procedure

4.3.2.1   Cell Preparation and Expression

Day 0 of procedure (about 16 to 18 h before starting Day 1):

1. Add 30 mL of LB media to a 125 mL baffled flask and sterilize. Add appropriate antibiotic if necessary following sterilization. Note: while you only need ~1 mL of culture per planned extract preparation for Day 1 of procedure, having additional liquid culture is suggested.

2. Inoculate media with desired *E. coli* strain from glycerol stock. Note: inoculating from a plate is also possible and should not make a substantial difference on lysate quality.

3. Incubate overnight at 37˚C, shaking at 250 rpm.

Day 1 of procedure:

4. Add 1 L of 2xYTPG media and no antibiotic to an autoclaved 4 L tunair flask. Note: we do not use antibiotics at this stage of the procedure. Antibiotics can reduce the lysate performance due to resulting cellular toxicity. 200 mL of cultures makes ~2 mL of extract.

5. Inoculate overnight culture (~1:100 dilution) into 2xYTPG media so that the starting OD_{600} is within the range of 0.05 to 0.1. Note: the 1:100 dilution is not a strict guideline but undergoing several doublings before induction is ideal.

6. Incubate culture at 37 ºC, shaking at 250 rpm.

7. Check the OD_{600} about every 30 minutes. Note: there can often be a variable lag phase, so it is important to check the OD_{600} frequently. Use 2xYTPG as a blank.

8. At OD_{600} of 0.5 to 0.8 (early exponential phase), induce recombinant protein overexpression by adding 0.1 mM (final concentration) IPTG. Note: IPTG concentrations can be tuned for optimal expression and other induction mechanisms can be used.

9. Incubate induced culture at 30 ºC, shaking at 250 rpm.

   - If enzyme overexpression is the objective, incubate in 30 ºC shaker for 4 h, or ~8 doublings, post-induction for protein expression. Note: This time can be optimized to obtain the desired weight of cells.

   - If cell-free protein synthesis is the objective, incubate in 30 ºC shaker to reach OD_{600} of 3. Check OD_{600} every ~30 min.

10. Following growth, immediately take the shake flask containing the culture out of the incubator and pour the contents into the pre-chilled 1 L centrifuge bottles on ice. Note: keeping cultures and subsequent handlings of the cultures on ice is imperative to maintain the quality of lysates produced.

11. Centrifuge samples in prechilled centrifuge bottles for 15 minutes at 5000 x g at 4 °C.

12. Wipe the inside walls of the centrifuge bottle with large Kim wipes to remove as much of the leftover supernatant as possible before inverting the bottle back upright. Note: removing as much supernatant at each centrifugation steps improves lysate quality.

13. Transfer cell pellet from centrifuge bottle to a pre-chilled 50 mL conical tube, using a spatula.

14. To obtain the remaining cells from the spatula and centrifuge bottle, add 2 mL of S30 buffer in the bottle and use the spatula to mix the cells into solution. Pipette the cell/buffer mixture into the 50 mL conical tube. Proceed with your remaining cell samples.

15. Place the 50 mL conical tube containing your cell sample submerged in ice.

16. Re-suspend in 25 mL S30 Buffer. Alternate between vortexing using a medium-high setting for 15 seconds and resting on ice for 15 seconds. Note: this can take 5 to 15 minutes or longer.

17. Once re-suspended, centrifuge samples in 50 mL conical tubes at 5,000xg for 10 min at 4 ºC to pellet cells.

18. Pour off supernatant into waste. Again, wipe the inside walls of the centrifuge bottle with Kim wipes to remove as much of the leftover supernatant as possible before inverting the bottle back upright.

19. Repeat steps 15 through 18.

20. Repeat step 16.

21. Once re-suspended, centrifuge samples in 50 mL conical tubes at 7,000xg for 10 min at 4 ºC to pellet cells. Note: the change is centrifugation speed is to keep the cells somewhat tight to remove residual liquid but somewhat loose before freeze-thaw to increase ease of resuspension.

22. Completely remove supernatant by pouring into waste. Before inverting your tubes back upright, clean the residual supernatant on inside of tubes as much as possible using a Kim wipe.

23. Use a clean pipette tip to split the cell pellet in half to increase the surface area of the pellet. Note: this will increase the speed of pellet resuspension when making extract.

24. Measure pellet mass on weight scale and record weight on tube.

25. Flash freeze the pellets in liquid nitrogen and store at -80 ºC. Note: Flash-freezing in liquid nitrogen helps maintain the quality of cells before extract preparation compared to just placing pellets in -80 ºC storage.

4.3.2.2   Extract Preparation

1. Take out cell pellets from -80 ºC freezer. Let pellets thaw on ice (~1 to 2 h).

2. Re-suspend pellets in 1 mL S30 Buffer per g of weighed pellet via vortex. It is important the cells are kept on ice. Vortex as necessary, but do not hold off ice for more than 15 seconds at a time. Note: this can take 15 min or more and the resuspension volume may be optimized for other strains.

3. Let samples rest on ice for suspension to settle.

4. Prepare a glass beaker with an ice water bath (90% ice, 10% water).

5. Transfer 1.4 mL of cell suspension into 1.5 mL micro-centrifuge tubes. Note: avoid transferring bubbles to 1.5 mL micro-centrifuge tubes.

6. Place micro-centrifuge tube containing cell suspension in ice bath.

7. Turn on the Q125 sonicator with 3.175 mm diameter probe at frequency of 20 kHz.

8. Place the sonicator tip into the cell suspension just below center of the micro-centrifuge tube. Note: tip should not rest against the sides of the tube. Moving tip around the cell suspension can improve energy transfer.

9. Set sonicator to 50% amplitude with pulsing at 10 seconds on, 10 seconds off. Note: this amplitude and pulsing time may be optimized for other strains.

10. Sonicate the 1.4 mL cell suspension with 820 J of sonication energy according to step 11 settings. Note: for a 1 mL suspension, the energy should be 530 J. For volumes between 1 mL and 1.4 mL, use linear interpolation to calculate sonication energy required. Refer to Kwon and Jewett for more information on sonication energies [169].

11. Different bacterial strains may require a pre-incubation step. Refer to Kwon and Jewett for more information on pre-incubation [169], noting that experiments have shown that variability at this step can lead to significant differences in extract performance. Note: for our previous pathway prototyping studies, pre-incubation was not needed.

12. Clean sonicator according to step 9 before moving to next sample. Repeat steps 9-13 for each sample.

13. Centrifuge samples at 12,000*g for 10min at 4 ºC.

14. Aliquot ~500 to 820 µL (if started with 1.4 mL extract mixture), top layer only, into a new tube. All samples from the original cell suspension should be pooled into the same new tube. Note: there may be 3 layers. Collect top layer only. In addition, the pellet after the first centrifugation tends to be loose. Therefore, immediately transfer the supernatant of the first centrifugation (typically with a pipette) into another tube to avoid carrying over layers. Experiments have shown that multiple centrifugations yield more active extracts as compared to extracts performed with one centrifugation.

15. Pipet mix thoroughly but gently. Aliquot 100 to 200 µL each (makes ~20 aliquots) on ice in new 1.5 mL or 0.6 mL micro-centrifuge tubes.

16. Flash freeze on liquid nitrogen and store at -80 ºC.

4.3.2.3   Extract Quantification of Total Protein by Bradford Assay

1. Remove the 1x dye reagent (QuickStart^{TM} Bradford Protein Assay Kit) and equilibrate to room temperature before use.

3. Refer to

4. **Table 4.2** for preparing the protein standards (use nanopure H_2O as the diluent) Note: we use the bovine serum albumin standard that comes with the QuickStart[TM] Bradford Protein Assay Kit.

Table 4.2. Bradford assay sample setup. *The following is the experimental setup of samples for extract quantification. This is adapted from the QuickStart*^{TM} *Bradford Protein Assay Kit (BioRad Laboratories, Inc.) instruction manual.*

Tube #	Source Volume (µL)	Source	Water Volume (µL)	[Protein] (µg/mL)
Standards				
1	10	2 mg/mL BSA stock	790	25
2	10	2 mg/mL BSA stock	990	20
3	6	2 mg/mL BSA stock	794	15
4	500	Tube 2	500	10
5	500	Tube 4	500	5
6	500	Tube 5	500	2.5
7	500	Tube 6	500	1.25
8	-	-	500	0
Extracts				
dilute extract	10	Extract to measure	990	
E4x	6	dilute extract	594	
E7x	8	dilute extract	552	
E10x	16	dilute extract	624	

5. Dilute extract samples (assuming extract concentration is ~40 mg/mL) by adding 10 µL extract to 990

    µL water. This will be the 'dilute extract' used to make further dilutions (

6. **Table 4.2**).

- Make a 10,000x dilution by adding 6 μL 'dilute extract' into 594 μL water.

- Make a 7,000x dilution by adding 8 μL 'dilute extract' into 552 μL water.

- Make a 4,000x dilution by adding 16 μL 'dilute extract' into 624 μL water.

5. Pipette 140 μL each standard and extract dilutions into separate microplate wells (refer to **Figure 4.3**).
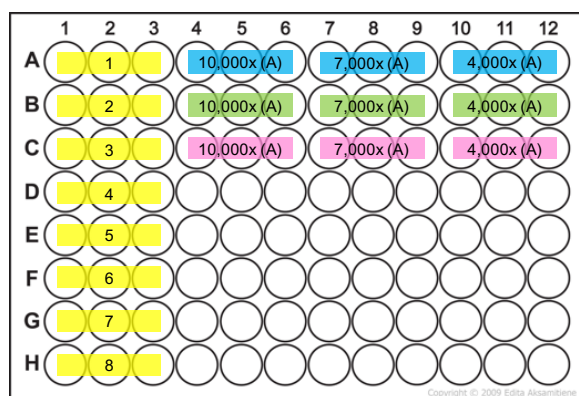
Figure 4.3. Layout of Bradford assay plate. Yellow (columns 1-3) are the standards. Blue, green, and purple (rows 1, 2, and 3; columns 4-12) are 3 separate extracts with triplicates for each of the 3 dilutions: 10,000x, 7,000x, and 4,000x.

6. Use a multichannel pipette to add 140 μL 1x dye reagent to each well and carefully mix. Note: avoid bubbles as they can drastically alter Bradford assay readings.

7. Incubate at room temperature for ~5 min. Samples should not be incubated longer than 1 h at room temperature.

8. Measure the absorbance of the standards and extract samples at 595 nm on a plate reader. Note: The linear range of these assays for BSA is 1.25–10 μg/mL.

9. Create a standard curve of the BSA standards measured and use this curve to calculate the protein content of each extraction dilution. Multiple each dilution to get a 1x concentration of the measurement. Then, average the values together to get an extract total protein concentration. Note: Total extract concentrations are typically between 35-45mg total *E. coli* protein/mL. Refer to QuickStart™ Bradford Protein Assay Kit instructions.

4.3.2.4   Overexpressed Protein Quantification by Densitometry

1. Add ~2 µL of extract to 3 µL NuPAGE® LDS Sample Buffer (4X), 0.6 µL 1M DTT, and water to make a total of 12 µL sample per extract being tested. Note: it is important for overexpressed protein quantification that a control sample of *E. coli* extract with no protein overexpressed is also made.

2. Heat each sample at 90 ºC for 10 min in a heat block.

3. Load 10 µL of each protein samples on a 4-12% Bis-Tris Nu-PAGE gel with 1x MOPS buffer and protein standards (SeeBlue plus2 ladder). Note: other protein ladders can be used in place of the SeeBlue plus2 ladder.

4. Run gel at a constant 120 V for 105 min.

5. After the run, open gel container and place gel in staining container.

6. Rinse 2x and shake in nanopure water for 5 min, then drain.

7. Stain with ~100 mL SimplyBlue™ SafeStain for 1 h on shaker. Note: any Coomassie stain can be used in place of SimplyBlue™ SafeStain

8. Rinse 2x and shake in water for 2 h or overnight.

9. Image on white background with Gel Doc™ XR+ Gel Documentation System using a Coomassie stain filter.

10. Once imaged the gel can be used to loosely quantify what percentage of protein in the extract is the overexpressed protein of interest using ImageJ software and documentation (http://rsb.info.nih.gov/ij/index.html). See **Figure 4.4** for details.
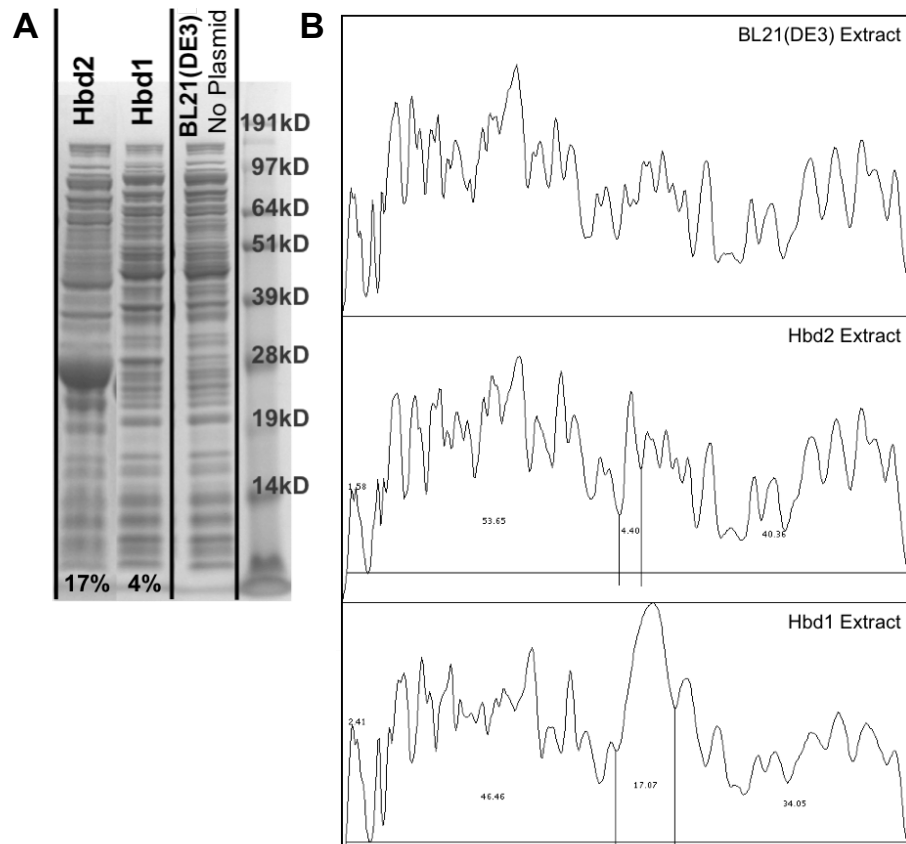
Figure 4.4. An example of semi-quantification overexpressed proteins in E. coli extracts. *(A) An SDS-PAGE gel is shown for 3 extracts: BL21 (DE3) containing no plasmid, containing a plasmid expressing Hbd1, and containing a plasmid expressing Hbd2. The percent of overexpressed protein relative to total protein content is listed at the bottom of each lane. (B) ImageJ analysis described is demonstrated here. This analysis is performed on the SDS-PAGE gel in A.*

4.4 Mix-and-match cell-free metabolic engineering

One approach to building pathways is to construct them by mixing lysates together that separately contain heterologous enzymes catalysing the chemical reactions to get to a product of interest. Rather than needing to engineer one organism to contain all enzymes of a pathway, we can overexpress single enzymes in chassis strains which can then be lysed and mixed to assemble the pathway. Provided are methods for biosynthetic reaction assembly.

4.4.1 Materials

4.4.1.1 Equipment

- 1.5 mL Eppendorf tubes

- 37 ºC incubator

4.4.1.2 Buffers and Reagents

- *E. coli* extracts individually expressing enzymes to assemble a biosynthetic pathway.

- 15x Salt Solution: magnesium glutamate (150 mM), ammonium glutamate (150 mM), and potassium glutamate (2010 mM).

- Glucose (2.2 M)

- Dipotassium phosphate (1 M, pH 7.2)

- Bis Tris (2 M)

- NAD (100 mM)

- ATP (100 mM)

- CoA (50 mM)

- DNAse/RNAse-free water

- 10%w/v trichloroacetic acid (TCA), sterile filtered.

4.4.2 Procedure

4.4.2.1 Mix-and-match biosynthesis reactions

1. Let reagents thaw on ice. Do not hand-thaw lysates.

2. For a first test of a pathway, label 21-1.5 mL Eppendorf tubes as 'reagent mix', 'extract mix', 'reaction mix', and 0, 3, 6, 9, 18, and 24 h time points in triplicate. It's important to do a time course for new pathways.

3. Keep all tubes on ice.

4. Assemble the 'reagent mix' by combining magnesium glutamate (8 mM), ammonium glutamate (10 mM), potassium glutamate (134 mM), glucose (200 mM), dipotassium phosphate (10 mM, pH 7.2), Bis Tris (100 mM), NAD (1 mM), ATP (1 mM), and CoA (1 mM). It is helpful to make this mix at 1.2x the

volume needed for all reaction tubes (18 tubes, 25 µL in this example). Mix thoroughly by vortex and keep tube on ice.

5. Assemble the 'extract mix' by combining each lysate pre-enriched with heterologous enzymes for the given pathway (in a five-enzyme pathway this would involve mixing five lysates). Start with each lysate at a final concentration of 2 mg/mL based on lysate quantification results. It is helpful to have this mix at 1.2x the volume needed for all reaction tubes. Gently pipette mix and keep on ice.

6. To create the 'reaction mix' combine the 'extract mix' and 'reagent mix' the volume in here should be at 1.1x the volume needed to pipette each reaction tube at 25 µL each. Keep on ice.

7. Pipette mix and transfer 25 µL of 'reaction mix' into each labelled reaction tube.

8. Immediately pull the 0 h time point (all replicates) and quench those reactions with 10% w/v trichloroacetic acid in a 1:1 ratio.

9. Incubate all other time points at 37 °C.

10. Pull each reaction with replicates at time points. Terminate reactions by adding 10% w/v trichloroacetic acid in a 1:1 ratio.

11. Precipitate proteins by pelleting through centrifugation at 15,000xg for 10 min.

12. The supernatant was stored at −80 ºC until analysis.

4.4.2.2   Biosynthesis Analysis

Metabolites can be quantified with current chromatography and mass spectroscopy techniques. In addition, chemical and enzymatic plate-based assays can be used when available [121,170,171,173].

4.5 Cell-free protein synthesis driven metabolic engineering

Using cell-free protein synthesis to enrich lysates with different enzymes for combinatorial assembly of different pathways enables parallelized pathway construction of combinatorial designs to accelerate DBT cycles.

4.5.1   Materials

4.5.1.1   Equipment

- 1.5 mL Eppendorf tubes

- 37 ºC incubator

- Microbeta scintillation detection instrument, PerkinElmer.

- Typhoon 7000 Imager (GE Healthcare Life Sciences, Pittsburgh, PA).

- Waterman chromatography paper.

- 12 in-square cellophane sheets.

- Autoradiography cassettes.

4.5.1.2   Buffers and Reagents

- *E. coli* BL21 Star(DE3) extracts with or without individually expressing enzymes as part of a biosynthetic pathway.

- DNA encoding each enzyme in the biosynthetic pathway.

- 15x Salt Solution: magnesium glutamate (150 mM), ammonium glutamate (150 mM), and potassium glutamate (2010 mM).

- 15x Nucleotide master mix: ATP (18 mM), GTP (12.75 mM), UTP (12.75 mM), CTP (12.75 mM), folinic acid (0.51 mg/mL), and 2.559 mg/mL tRNA.

- Amino acid solution containing all 20 amino acids at 50 mM.

- Phosphoenolpyruvate (PEP) at 1M.

- Putrescine (250 mM).

- Spermidine (250 mM).

- HEPES Buffer (1 M, pH 7.2).

- Glucose (2.2 M)

- Dipotassium phosphate (1 M, pH 7.2)

- NAD (100 mM)

- ATP (100 mM)

- CoA (50 mM)

- DNAse/RNAse-free water

- ^{14}C-Leucine (10 µM)

- 0.1 M sodium hydroxide.

- 5%w/v trichloroacetic acid (TCA).

- 10%w/v trichloroacetic acid (TCA), sterile filtered.

- Cytoscint™-ES liquid scintillation cocktail.

- SplitGFP Buffer: 50 mM Tris pH 7.4, 0.1 M NaCl, 10% glycerol (TNG buffer) [184].

4.5.2　Procedure

CFPS reactions were performed to express enzymes involved in *n*-butanol production prior to starting the CFME portion of the reactions using a modified PANOx-SP system [152,153].

4.5.2.1　CFPS-ME Reactions

1. Let reagents thaw on ice. Do not hand-thaw lysates.

2. For a first test of a pathway, label 21-1.5 mL Eppendorf tubes as 'reagent mix', 'extract mix', 'reaction mix', and 0, 3, 6, 9, 18, and 24 h time points in triplicate. It's important to do a time course for new pathways.

3. Keep all tubes on ice.

4. Assemble the 'CFPS Master Mix' by combining 15x Salt Solution (magnesium glutamate (8 mM); ammonium glutamate (10 mM); potassium glutamate (134 mM)), 15x Nucleotide Mix (ATP (1.2 mM); GTP, UTP, and CTP (0.85 mM each); folinic acid (34.0 µg/mL); *E. coli* tRNA mixture (170.0 µg/mL)), 20 standard amino acids (2 mM each), NAD (0.33 mm), CoA (0.27 mM), spermidine (1.5 mM), putrescine (1 mM), and PEP (33 mM). It is helpful to make this mix at 1.2x the volume needed for all reaction tubes (18 tubes, 22.5 µL in this example). Mix thoroughly by vortex and keep tube on ice.

5. Assemble the 'extract mix' by using plain *E. coli* extract with no enzyme enrichment or by combining each lysate pre-enriched with heterologous enzymes for the given pathway (in a five-enzyme pathway this would involve mixing five lysates). Start with each lysate at a final concentration of 2 mg/mL based

on lysate quantification results. The total lysate concentration should be 10 mg/mL. It is helpful to have this mix at 1.2x the volume needed for all reaction tubes. Gently pipette mix and keep on ice.

6. To create the 'reaction mix' combine the 'extract mix', 'reagent mix', and plasmid DNA encoding each enzyme (~13.3 ng/mL each). The volume in here should be at 1.1x the volume needed to pipette each reaction tube at 22.5 µL each. Keep on ice.

7. Pipette mix and transfer 22.5 µL of 'reaction mix' into each labelled reaction tube. Note: it is important that the reaction components are evenly mixed. Cell-free protein synthesis has been plagued by variable results. One cause for this variability in our laboratory has been the lack of complete mixing of the reaction components. Specifically, this can be a problem when setting up numerous reactions from the same pre-mix. Air bubbles, especially in the extract can also be an issue.

8. Incubate all reaction tubes at 30 °C.

9. Assemble the 'CFME Master Mix' by combining glucose (200 mM), NAD (0.67 mM), and CoA (0.73 mM). Mix thoroughly by vortex and keep tube on ice.

10. After 3 h, spike in 2.5 µL 'CFME Master Mix' to each reaction tube to initiate biosynthetic activity.

11. Immediately pull the '0 h' time point (all replicates) and quench those reactions with 10% w/v trichloroacetic acid in a 1:1 ratio.

12. Incubate all other time points at 30 °C.

13. Pull each reaction with replicates at time points. Terminate reactions by adding 10% w/v trichloroacetic acid in a 1:1 ratio.

4.5.2.2   Quantification of protein produced in vitro via radioactive incorporation

1. Using the procedure for CFPS-ME Reactions in section 5.2.1, the 'CFPS Master Mix' can be made with radioactive ^{14}C-Leucine (10 µM) at step 4.

2. Samples are quenched with 100 µL of 0.1 M sodium hydroxide and incubated at 37 ºC for 20 min.

3. Quenched samples are then split in half ('washed' vs. 'unwashed') and pipetted onto 0.25-in by 1-in Waterman paper strips. The strips are then dried.

4. The 'washed' half of samples are then washed three times with 5%w/v TCA to precipitate radioactive protein samples.

5. The 'washed' half of samples are then washed with molecular biology grade ethanol.

6. Radioactivity of TCA-precipitated samples and 'unwashed' were measured by liquid scintillation counting to then quantify the protein produced as previously reported (MicroBeta2; PerkinElmer)[51,153].

7. These reactions were also run on a Coomassie-stained SDS-PAGE gel as done in section 3.2.4 steps 1 through 11.

8. The gels can then be dried between 2-12 in square cellophane sheets overnight.

9. The dried gel can now be exposed by autoradiography for 3 days.

10. Autoradiographs are imaged with a Typhoon 7000 Imager (GE Healthcare Life Sciences, Pittsburgh, PA).

11. Multiple proteins produced *in vitro* were further quantified by gel image intensity comparisons using ImageJ (NIH) similar to section 3.2.4 (**Figure 4.5A**).
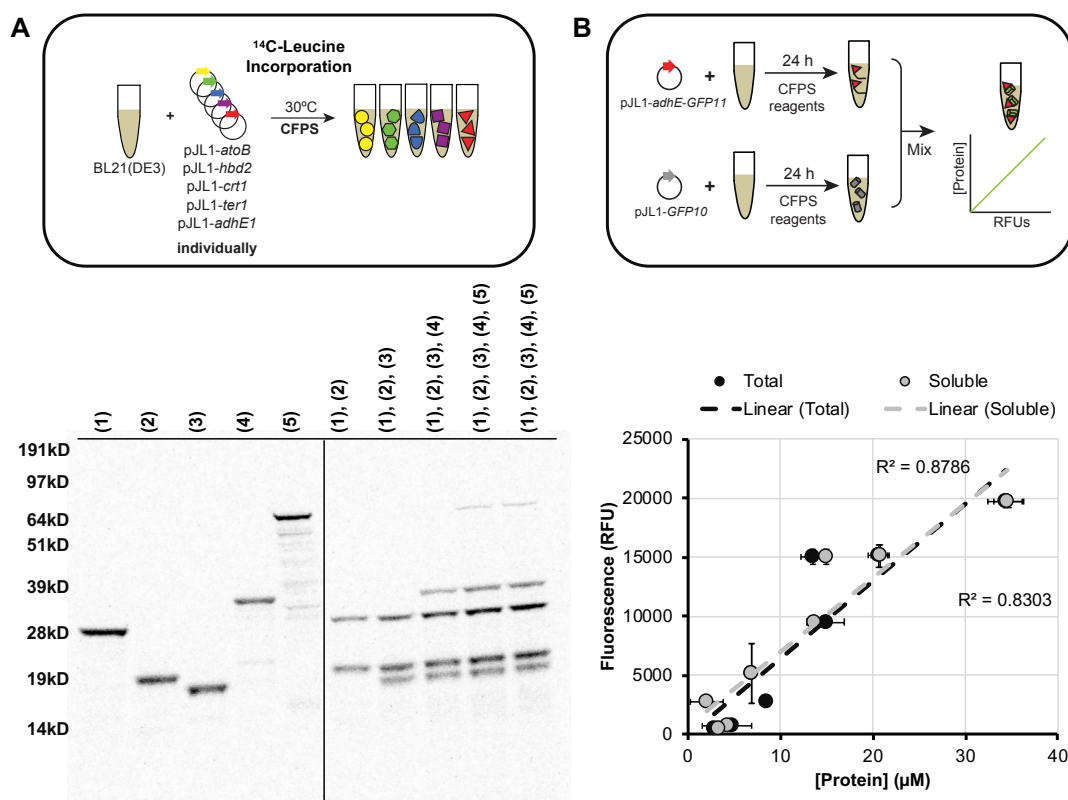
Figure 4.5. Quantification of cell-free protein synthesis examples. *(A) represents quantification by radioactivity. SDS-PAGE autoradiogram shows each protein produced by cell-free protein synthesis. Each column is a separate cell-free reaction producing one of five proteins or a combination of the five proteins. (B) represents quantification by splitGFP reporter. Each circle represents a concentration of protein measured by radioactive* ^{14}*C-leucine incorporation and the corresponding fluorescent readout.*

4.5.2.3   Quantification of protein produced in vitro via split-GFP construct

If radioactive ^{14}C-Leucine is unavailable, other quantification methods can be used. Here we present a method for quantification by split-GFP fluorescence. By adding a 20-amino acid 'GFP11' tag onto the end of each protein and expressing the corresponding 'GFP1-10' protein, the association of the two will elicit a quantifiable fluorescent signal.

1. DNA constructs encoding enzymes to prototype can be designed to include the following encoded amino acid sequence directly at the end of the coding sequence before the stop codon 'DGGSGGGSTSRDHMVLHEYV'.

2. Also, a DNA construct pJL1-splitGFP encoding the other portion of GFP can be made 'MGGTSMSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFICTTGKLPVPWPTLVT TLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTD FKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGDGPVLLP DNHYLSTQTVLSKDPNEK' [184].

3. Perform CFPS reactions of pJL1-gene-gfp according to section 5.2.1, steps 1 through 9 (CFPS RXN 1).

4. Separately, perform CFPS reaction of pJL1-splitGFP (CFPS RXN 2).

5. Mix 10 µL of CFPS RXN 1, 5 µL of CFPS RXN 2, and 5 µL SplitGFP Buffer.

6. Measure fluorescence over a 20 h period (**Figure 4.5B**).

4.5.2.4   Metabolite Quantification

Metabolites can be quantified with current chromatography and mass spectroscopy techniques [170,174]. In addition, chemical and enzymatic plate-based assays can be used when available through commercially available kits or documented in the literature.

4.6 Summary

Cell-free synthetic biology provides powerful tools to prototype biosynthetic pathways, providing an unprecedented capability to test hundreds to thousands of pathways by avoiding inherent limitations of cell growth. The methods provided here describe multiple ways of constructing cell-free enzymatic pathways to iterate through DBT cycles at speeds 10x faster than traditional approaches. Coupling cell-free protein synthesis, in particular, to the construction of a metabolic pathway in tandem with high-end metabolomics will offer a high degree of flexibility to model the kinetics and stability of individual enzymes, measure metabolite fluxes in multistep pathways, and experimentally isolate many other parameters confounded in living organisms. Overall, the modular and cell-free nature of our framework is poised to facilitate multiplexed, automated study of biosynthetic pathways to inform systems-level design.

4.7 Conclusions

This chapter detailed the methods to build all cell-free systems I designed in my PhD. My hope is that this framework of building pathways in cell-free systems will be widely used. By providing detailed instructions my hope will more likely become a reality. While this provides an odd chapter for a dissertation I believe it provides an essential one. Other scientists must be able to reproduce methods for them to be assimilated into a modern repertoire of synthetic biology techniques.

Chapter 5: Automation for cell-free prototyping

If you want to make an apple pie from scratch, you must first create the universe.

- Carl Sagan

This chapter is the detailed optimization of a cell-free system in a new chassis strain using liquid-handling robotics and machine-learning. While the work does not deal with biosynthetic pathways, it does involve new metabolic schemes. My main contribution to this work was in developing the liquid-handling robotics to accomplish this optimization. The system itself is the in vitro ribosome synthesis, assembly, and translation system developed in the lab, and provided a greater-than-24-component system to optimize. This work provides precedent for the liquid-handling robotics and physiochemical optimizations of Chapter 6 and the machine-learning implemented in Chapter 7.

5.1 Abstract

Building variant ribosomes offers opportunities to reveal fundamental principles underlying ribosome biogenesis and to make ribosomes with altered properties. However, cell viability limits mutations that can be made to the ribosome. To address this limitation, the *in vitro* integrated synthesis, assembly and translation (iSAT) method for ribosome construction from the bottom up was recently developed. Unfortunately, iSAT is complex, costly, and laborious to researchers, partially due to the high cost of reaction buffer containing over 20 components. In this study, we develop iSAT in *E. coli* BL21Rosetta2, a commonly used bacterial strain, with a cost-effective poly-sugar and nucleotides monophosphate-based metabolic scheme. We achieved a ten-fold increase in protein yield over our base case with an evolutionary

design of experiments approach, screening over 490 reaction conditions to optimize the reaction buffer. The computationally-guided, cell-free, high-throughput technology presented here augments the way we approach multi-component synthetic biology projects and efforts to repurpose ribosomes.

5.2 Introduction

The bacterial ribosome is a macromolecular machine evolutionarily optimized for the template-guided, sequence-defined polymerization of amino acids into proteins. By building ribosomes, synthetic biology efforts seek to elucidate a new understanding of the science of protein synthesis through creation – in the sense of Feynman's dictum, "What I cannot create, I do not understand." These efforts aim to reveal fundamental principles that underlie the operation and assembly of the ribosome complex and translation,[185-187] design and build minimal cells to understand the origins of life,[188,189] and enable evolution to select for ribosomes that have enhanced functions, such as altered chemical and functional properties.[190-192] However, manipulation of ribosomes in bacteria is often limited by cell viability constraints. In vitro assembly, or reconstruction of E. coli ribosomes from purified native ribosomal components into functionally active 30S and 50S ribosomal subunits, is a promising alternative to study the ribosome.[193-195] Until recently, however, *in vitro* assembly of E. coli ribosomes has been limited because conventional ribosome reconstitutions are non-physiological, and ribosomes reconstituted with in vitro transcribed ribosomal RNA (rRNA) are essentially non-functional.[196,197]

To address the limitation, our lab has developed over the course of the last several years, the *in vitro* integrated synthesis, assembly and translation (iSAT) method for ribosome construction.[198-201] In a single reaction, the iSAT method constructs ribosomes in a cell-free, ribosome-free (S150) extract by transcribing DNA encoding ribosomal RNA (rRNA), and then processing and assembling transcribed rRNA and ribosomal proteins (r-proteins) into ribosomes that translate reporter proteins. Recent work on iSAT has dramatically improved the platform by optimizing extract preparation methods,[200] tuning rRNA

transcription,[199] identifying and alleviating substrate limitations,[201] and using macromolecular crowding and reducing agents.[198]

The iSAT system is a unique platform, which could be potentially used for bottom-up construction of minimal cells.[189,202,203] In this context, we recently demonstrated the ability to build functionally active ribosomes using iSAT in giant liposomes. The liposomes were prepared using double emulsion template, and compartmentalized *in vitro* protein synthesis was analyzed using spinning disk confocal microscopy.[204] This was the first time that a cell-free transcription and translation system where the DNA molecule encoding the formation of ribosomes has been encapsulated in a model cell-like compartment, *i.e.* liposome. While this was an important step towards the construction of minimal cells, iSAT is still complex, costly and laborious to researchers, partially due to the high cost of reaction buffer containing over 20 components. To make the system more accessible, we hypothesized that the price per reaction could be decreased by using different chassis for lysate preparation with different energy regeneration metabolisms, yet it is unclear if metabolic enzymes present in the S150 extracts used for iSAT are sufficient to support ribosome synthesis, assembly and translation.

Historically, cell-free systems have been complex and expensive molecular mixtures, owing to many different chemicals and high-energy phosphate compounds that drive energy regeneration[51,152,205,206]. Glucose, PEP, pyruvate, 3-PGA, cellulose, etc. have all been evaluated for CFPS with some promising results. Recently, a high yielding, cost-effective metabolic scheme[207,208] was developed in lysates made from *E. coli* BL21Rosetta2, a common *in vivo* production strain, which could potential bring the cost of iSAT down. However, the traditional iSAT system has been developed with *E. coli* MRE600, a common strain for ribosome study,[200] due to its low RNase I activity. We wondered if we could activate iSAT with a simplified and cost-effective energy metabolism in lysates from the *E. coli* BL21Rosetta2, a B strain phylogenetically divergent from MRE600[209].

We therefore set out to: (1) develop the iSAT system in *E. coli* BL21Rosetta2, a more commonly used bacterial strain, (2) create a simplified and cost-effective, poly-sugar and nucleotides monophosphate-based metabolic scheme that fuels iSAT, and (3) optimize the reaction buffer for cell-free protein synthesis in these lysates through high-throughput, combinatorial optimization over the 20 experimental components described in **Table 5.1**. A key challenge for our goal is that finding an effective experimental design for the

optimization of a 20-dimensional experimental space is hard. Conventional approaches to the design-of-experiments problem typically aim at reducing the number of experimental parameters (i.e., components) to explore, in order to make the exhaustive search of the resulting lower-dimensional experimental space feasible. Our approach consists, instead, of a form of Evolutionary Design-of-Experiments (EDoE),[210-212] where predictive modeling and artificial intelligence guide the iterative selection of small batches or "generations" of experiments to perform, with each generation corresponding to a different small sample of the actual full-dimensional space (**Figure 5.1**). This approach also allows us to make observations of the complex chemical interactions occurring in the cell-free system.

Table 5.1. The 20 components included in iSAT reaction buffer tested for E. coli BL21Rosetta2 lysates. *Each of the 20 components are placed in one of 8 categories. Four concentrations of each component (Level 1 through 4) are represented here as varied concentration. These are used to change each component individually in the experimental setup. The pivots of all generations are also shown. The concentration of each component for a given pivot is listed in each generation's column. These values are fixed for 19 components with the remaining component being varied at the specified varied concentrations.*

	Category	Component	Unit	Varied Concentration				Pivot Generation							
				Level 1	Level 2	Level 3	Level 4	1.1	1.2	2	3	4	5	6	7
1	Phosphate donor	AMP	mM	0	0.65	1.3	7.3	1.3	1.3	0	0.65	0	0	0	0
2	Phosphate donor	UMP	mM	0	0.41	0.82	4.59	0.82	0.41	4.59	0.82	0.82	0.41	0	4.59
3	Phosphate donor	CMP	mM	0	0.41	0.82	4.59	0.82	4.59	0.41	0.82	0.41	0	0	0.82
4	Phosphate donor	GTP	mM	0	0.72	1.43	8.03	1.43	0.72	8.03	8.03	8.03	8.03	8.03	8.03
5	Phosphate donor	Acetyl-phosphate	mM	0	0.16	0.32	1.79	0.32	0	0.16	0.16	0.32	0.32	0.32	0.16
6	Phosphate donor	K2HPO4/KH2PO4	mM	0	2.41	4.82	27.02	4.82	4.82	4.82	2.41	4.82	2.41	2.41	0
7	iP recycling	Maltodextrin	mM	0	16.26	29.8	44.16	2.53	42.56	29.8	29.8	44.16	44.16	29.8	16.26
8	Energy Source	3-PGA/PEP 1:1.7	mM	0	6.24	12.48	69.87	12.48	0	69.87	69.87	69.87	0	0	0
9	Cytomim	Folinic Acid	µg/mL	0	14.8	29.59	165.71	29.59	165.71	14.8	29.59	29.59	0	14.8	14.8
10	Cytomim	Putrescine	mM	0	0.92	1.85	10.34	1.85	0	0.92	0.92	1.85	1.85	0	0
11	Cytomim	Spermidine	mM	0	0.64	1.28	7.14	1.28	7.14	1.28	7.14	1.28	1.28	0	0
12	Salt	K-Glu	mM	0	14.46	200	250	8.93	250	0	200	14.46	14.46	14.46	0
13	Salt	K-OX	mM	0	1.28	2.55	14.29	2.55	1.28	2.55	1.28	0	0	0	0
14	Salt	Mg-Glu	mM	0	3.28	6.56	36.72	6.56	36.72	3.28	6.56	6.56	3.28	6.56	0
15	Translation	tRNA	µg/mL	0	82.02	164.03	918.57	164.03	0	82.02	164.03	0	164.03	0	0
16	Translation	Amino acids	mM	0	1.04	2.08	11.64	2.08	11.64	1.04	1.04	2.08	1.04	0	0
17	Cofactor	cAMP	mM	0	0.33	0.66	3.7	0.66	3.7	3.7	0.33	0.66	0.66	0	0
18	Cofactor	CoA	mM	0	0.13	0.25	1.41	0.25	1.41	0.13	0.25	0	0	0	0
19	Cofactor	NAD	mM	0	0.16	0.32	1.79	0.32	0	0.16	0	1.79	1.79	1.79	1.79
20	Reducing agent	DTBA	mM	0	0.96	1.93	10.80	1.93	10.8	1.93	1.93	0	0.96	10.8	1.93
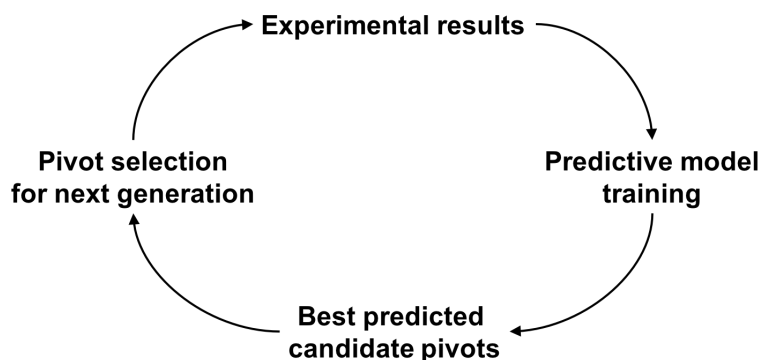
Figure 5.1. Evolutionary Design of Experiments Approach. *Illustration of the Evolutionary Design of Experiments process, iteratively performing each generation of experiments, beginning with that generation's experimental results. The design of each generation is based on a predictive model built on data from all generations, including the current one. Multiple candidate pivot experiments are selected based on the model's predictions and one of them is chosen by the experimentalist for programming the liquid handling robot.*

In this work, we use EDoE to optimally implement a complex metabolic scheme to fuel the iSAT molecular interactions (**Figure 5.2**). The experimental design is based on a library of small molecules used to revitalize endogenous enzymes already present in the S150 cytoplasmic extract[200] which includes a new poly-sugar metabolic scheme. The new metabolic scheme was designed taking inspiration from an efficient and cost-effective poly-sugar based metabolism for ATP-regeneration,[207,208,213] and nucleotides triphosphate regeneration.[214] The molecular composition of this novel metabolic scheme has not been described before. We exploited the power of liquid-handling robotics to build cell-free iSAT reactions, which can then be tested for reporter protein translation and optimized via EDoE. With this powerful approach, we could overcome a barrier of complexity given by the many molecular component interactions involved in the ribosome assembly and protein synthesis *in vitro*, which simplifies reaction set-up. We achieved a ten-fold increase of protein yield of our base case with our EDoE approach. The computationally-guided, cell-free, high-throughput technology presented here alters the way we can approach complex, multi-component synthetic biology projects, providing a path forward for improving cell-free efforts in: *in vitro* ribosome engineering,[215] minimal cell synthesis,[202] quorum sensing,[216] gene circuits optimization,[217] metabolic engineering,[144,173] biocatalyst discovery,[218] directed evolution,[219] glycosylation,[220] non-standard amino acid incorporation,[221,222] and human protein synthesis.[223]
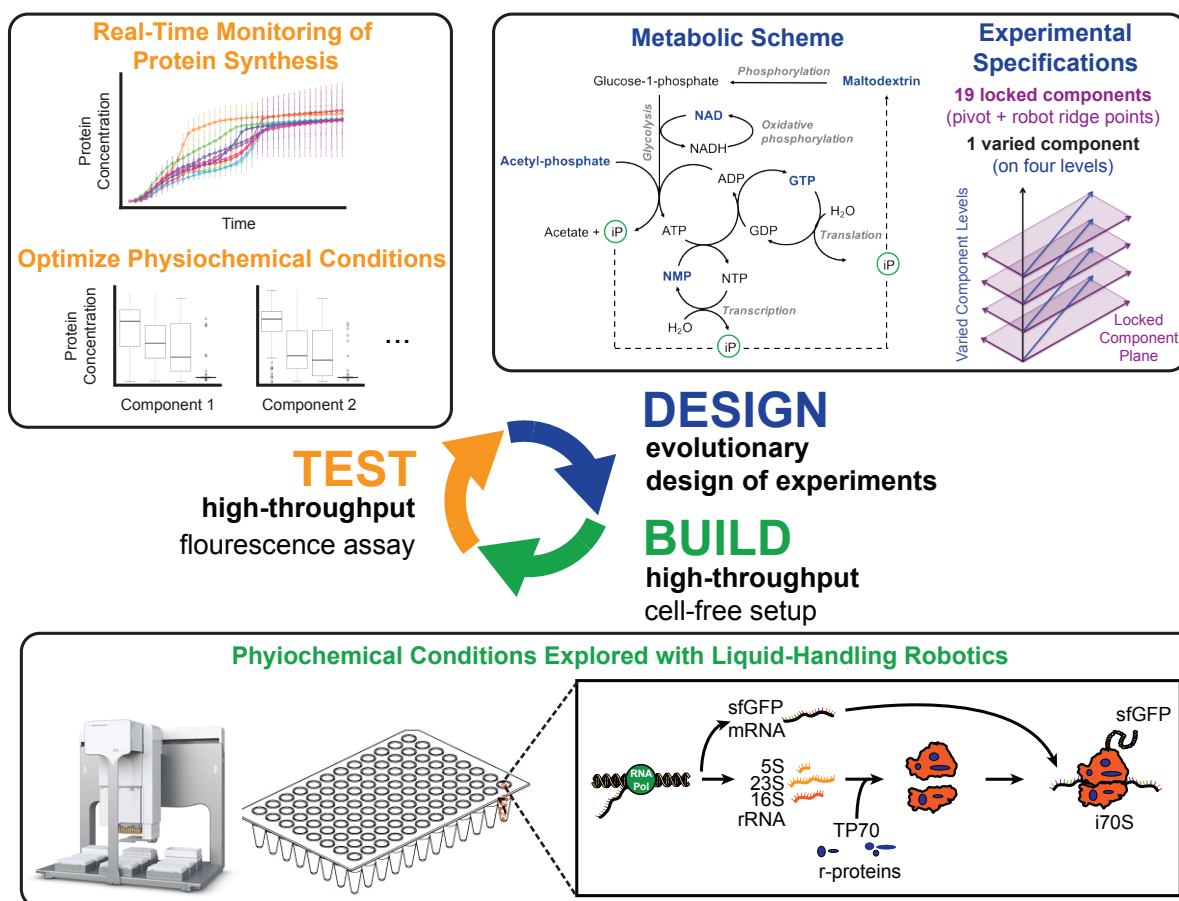
Figure 5.2. Overview schematic of iterative Design-Build-Test platform for complex metabolic optimizations.

5.3 Materials and Methods

5.3.1 Bacterial strains and plasmids

E. coli BL21Rosetta2 (NEB) was used in creating S150 cell lysates for use in all iSAT experimental generations. Plasmid pJL1-sfGFP was used for reporter expression, and plasmid pT7AM552A was used for rRNA operon expression.

5.3.2 S150 cell extract and component preparation

E. coli BL21Rosetta2 cells were grown in 2x YTP media (16 g l^{-1} tryptone, 10 g l^{-1} yeast extract, 5 g l^{-1} NaCl, 7 g l^{-1} potassium phosphate monobasic, 3 g l^{-1} potassium phosphate dibasic). S150 extract, *E.*

coli 70S ribosomes, total protein of 70S ribosomes (TP70) and T7 RNA polymerase (RNAP) were prepared as previously reported.[199] The amino acids mixtures used by the robotic workstation were prepared as previously described.[224]

5.3.3   iSAT reaction setup

The typical iSAT system is composed of 22 components listed in **Table 5.2** with traditional concentrations. In the new metabolic scheme explored for *E. coli* BL21Rosetta2 lysates, the iSAT reaction was composed of *E. coli* S150 crude extract (3.7 µg/mL), rRNA operon plasmid (pT7AM552A, 4.0 nM), pJL1-sfGFP (4.0 nM), T7 RNA polymerase (36 µg/mL), total protein of 70S ribosomes (TP70, 300 nM), and the 20 different components listed in **Table 5.1**. These 20 components, plus the homemade amino acids mixture, make up the reaction buffer necessary for fueling *in vitro* protein synthesis.

Table 5.2. Final concentrations of components used in traditional iSAT reactions.

Reagent	Traditional Concentration	
Magnesium glutamate (Mg(Glu)$_2$)	14	mM
Ammonium glutamate (NH$_4$(Glu))	10	mM
Potassium glutamate (KGlu)	265	mM
Spermidine	2.0	mM
Putrescine	1.5	mM
ATP	1.20	mM
GTP	0.85	mM
UTP	0.85	mM
CTP	0.85	mM
Folinic Acid	34.0	µg/mL
tRNA	171	µg/mL
20 amino acids	2.00	mM
NAD	0.33	mM
CoA	0.27	mM
HEPES-KOH, pH 7.6 (total in reaction)	80.00	mM
Oxalic acid	4.00	mM
PEP	42.00	mM
E. coli S150 crude extract (MRE600)	3.7	µg/mL
Operon plasmid: pT7AM552A	4.0	nM
pJL1-sfGFP	4.0	nM
T7 RNA polymerase	36	µg/mL
Total protein of 70S ribosomes (TP70)	300	nM

5.3.4 Robotic liquid-handling reaction setup and exploration of the experimental space using pivots

An Agilent BRAVO liquid-handling robot (Agilent Technologies) was used to carry out the iSAT reaction buffer setup. The liquid-handling workstation has nine plate decks and a 96-pipette tip head movable in the x-y-z directions. The workstation was programmed using VWorks™ Automation Control Software (BioNex Solutions, Inc.) to pipette different arrangements of reagents. Concentration levels for each component was made through serial dilution of stock vials of each of the 20 components using the liquid-handling system (both the 96ST and 96LT heads). Each component was diluted with 50 mM HEPES pH 7.2 buffer. After initial component dilutions are made, the reagents are pooled to together in 96-well flat-bottom plates. The exploration compatible with our robotic constraints is conducted by choosing a "pivot" experiment, and then performing alterations on the pivot. The alterations are defined by keeping 19 components fixed at the pivot concentrations (**Table 5.3**) while the concentration of the remaining component is varied one level at the time (the concentration levels are highlighted using a color code in **Table 5.1**). For example, well A1 contains all reagents except AMP, and A1 through A4 contain AMP at each of the four concentration levels (low, medium-low, medium-high, and high) listed in **Table 5.1**; these complete reaction buffers are then mixed with S150 extract, purified ribosomal proteins, DNA, etc. to complete the iSAT reaction mixture. The set of all single-component alterations of a pivot make up the pivot "neighborhood". The pivot neighborhood is comprised of 60 different experiments since each component may be varied to three alternative values besides that component's pivot value. The generation of experiments determined by a particular pivot contains all the experiments in the pivot neighborhood, and in addition, 20 replicates of the pivot experiment, yielding 80 experiments fitting on a 96-well plate. Reactions are run in duplicate, and sfGFP fluorescence was measured at 37°C in sealed 500 µL tubes (Biorad) using a real time PCR machine.

5.3.5 GFP Quantification

The kinetic data and yields of sfGFP were measured through excitation at 485 nm while measuring emission at 528 nm with a 510 nm cutoff filter. The fluorescence response of sfGFP was converted to concentration according to a standard curve as previously described.[162]

5.3.6 Modeling and design of experiments

The use of modeling in the experimental process is illustrated in **Figure 5.1** and **Figure 5.2**. Each generation begins with the collection of experimental results for that generation. The experiments for the following generation are designed by first building a predictive model from the data for all completed generations, including the current one. The predictive model takes an experiment as input, and outputs a prediction for the experimental result expected for that experiment, i.e., a value on the modeled experimental response surface. The predictive model is used to explore the experimental space, through a process of "virtual execution" of many randomly sampled pivots, combined with hill-climbing of the model response surface, as described below. The trade-off between random exploration and hill-climbing is varied each generation as described in the detail below, but at the end of each generation's exploratory process four good candidate pivots are automatically chosen, and the experimentalist selects between these four (on the basis of intuition combined with expert knowledge). The pivot selected through this process is then used to determine the following generation, as described above.

The model constructed has the form of an ensemble of 25 single-hidden-layer, feed-forward neural networks, each network initialized with different random initial weights and trained via the *nnet* package in R. Each network has 20 input nodes, corresponding to the 20 dimensions (components) of the experimental space, and one output node, to predict the experimental response given values for all inputs. Prior to training, the 20 replicates of each pivot experiment are collapsed into one individual data point, whose response value is the mean of the response value across the 20 replicates. After training, a network may be used to predict experimental response for an arbitrary set of inputs, i.e. for any candidate experiment. The ensemble prediction is obtained by taking the mean of the predictions of each of the 25 networks in the ensemble.

Predictive models must be carefully controlled by a process of regularization, to avoid overfitting. We regularized our network ensemble models by exploring a range of model hyper-parameters in a bootstrapping process. The hyper-parameters explored were the number of hidden nodes for the networks (values 2, 5, 10, 20, 40), the value of the weight-decay term (values 0.01, 0.05, 0.1, 0.25, 0.5, 1), and the number of iterations of the back-propagation training algorithm (values 100, 500, 1000, 2000). For each

combination of these hyper-parameters, the trained network ensemble was bootstrapped via Monte-Carlo cross-validation, on 20 independent random partitions of the data from all experiments and measured responses up to the last completed generation into training and cross-validation sets (80% training / 20% validation). Each hyper-parameter combination was assigned a score corresponding to the mean across partitions of the Pearson's correlation between predicted and observed response values in the validation set.

Once the hyper-parameter combination with the highest score has been selected through the cross-validation process, a predictive model with such hyper-parameter values is trained on the entire data from all experiments and measured responses up to the last completed generation. Given the predictive model, any given candidate pivot may be evaluated by computing its predicted experimental response as well as that for each experiment in its neighborhood. The predicted response associated to that pivot is then taken to be the 90% quantile of the distribution of predicted responses for the set of experiments including the pivot and its neighborhood.

Given a predictive model built from the selected hyper-parameters, we can proceed to choosing good predicted experiments as candidate pivots for the next generation. These candidate pivots are chosen by combining random sampling with two techniques that manage the trade-off between random exploration and exploitation of information gathered in previous generations: (i) a steepest-ascent algorithm to hill-climb the predicted response surface to reach local maxima (ii) an algorithm to constrain experiment choice based on "experimental distance", defined as the mean pairwise Euclidean distance between each experiment in the set that includes the candidate pivot and its neighborhood and each experiment in the set of already performed experiments, calculated after normalizing the range of each component to a [0, 1] scale. The details of the experimental choice process vary from generation to generation, based on the complexity of the predicted response surface, and the evaluation of experimental results for each generation. A summary of the experimental process for different generations follows:

- Generation 1.1-1.2: chosen by the experimental team.
- Generation 2-3: randomly sample 250 candidate pivots, then select the four pivots with best predicted response among those having experimental distances falling between the median and the 3rd quartile of the distribution. The experimental team finally chooses one from these four.

- Generation 4: Run the steepest-ascent algorithm from 500 randomly sampled initial pivots; then randomly sample four candidate pivots from the local maxima with predicted response above the median of the distribution and experimental distance falling between the median and the 3rd quartile of the distribution. The experimental team finally chooses one from these four.

- Generation 5: Run the steepest-ascent algorithm from 5000 randomly sampled initial pivots; randomly sample 250 pivots from the local maxima with predicted response above the median of the distribution, with probabilities proportional to their respective predicted response; then select the four candidate pivots with best predicted response among those having experimental distance falling between the median and the 3rd quartile of the distribution. The experimental team finally chooses one from these four.

- Generation 6: Run the steepest-ascent algorithm from 20000 randomly sampled initial pivots; randomly sample 250 pivots from the local maxima with predicted response above the 3rd quartile of the distribution, with probabilities proportional to their respective predicted response; then select the four candidate pivots with best predicted response among those having experimental distance falling between the 1st and the 3rd quartile of the distribution. The experimental team finally chooses one from these four.

- Generation 7: Run the steepest-ascent algorithm from 80000 randomly sampled initial pivots; randomly select 250 pivots from the local maxima with predicted response above the 90% quantile of the distribution, with probabilities proportional to their respective predicted response; then select the four candidate pivots with best predicted response. The experimental team finally chooses one from these four.

There was one experiment in generation 2 that appeared to have extremely high response; subsequent experiments have not validated its repeatability, so that experiment has been deleted from this presentation of results. The data point corresponding to this severely noisy experiment was, however, used in the model training process, causing bias in the predictions for generation 3.

5.4   Results

The goal for this work was to develop a robust, cost-efficient iSAT system by studying the interactions of molecular components present in during iSAT reactions. Specifically, we wanted to develop iSAT in the commonly-used *E. coli* BL21Rosetta2 strain with a poly-sugar substrate, taking advantage of nucleotide recycling. To achieve this goal, we used an evolutionary design of experiments (EDoE) approach, and liquid- handling robotics, to efficiently explore the interactions between each of the iSAT system components and carry out several rounds of optimization. This allowed us to develop and improve iSAT in BL21Rosetta2 with a cost-efficient energy metabolism.

5.4.1   Design of the experimental space

We wanted to develop a poly-sugar metabolic scheme with the iSAT system in E. coli BL21 Rosetta2 to make iSAT simpler and cost-effective. This metabolic scheme was chosen in order to activate complex cell-free metabolism and has been shown to work in other cell-free systems.[43] This scheme consists of the synthesis of nucleotides triphosphates from nucleotide monophosphates coupled to glycolysis activation upon hydrolysis of maltodextrin from inorganic phosphate, which is the byproduct of cell-free protein synthesis.[207,208] This design bypasses an energy re-generation system using substrate level phosphorylation, which is important to reduce the overall cost of the cell-free reaction,[208] but also in the design and integration of the sub-systems of minimal cells (i.e., container, information, and metabolism), an important related research area.[202]

To activate such a metabolic scheme, we investigated 20 potentially beneficial small molecules. We chose to create design space that covered these 20 components subdivided into 8 categories, each one with a specific function needed for ribosome assembly and protein synthesis (**Table 5.1**). The first group of components is the phosphate donor category, which contains the molecules that upon hydrolysis release inorganic phosphate (iP) in the reaction environment. To follow, maltodextrin is considered the iP recycling element, incorporating the phosphate in glucose-1-phosphate, and activates glycolytic pathways in vitro for ATP synthesis and regeneration.[202,207,208] A mixture of 3-PGA and PEP was investigated as the energy source for ATP synthesis.[205,225] In addition, molecules of the Cytomim energy regeneration system

were included in the screening process as well.[51,152] Moreover, typical components such as: salts, co-factors and purified ribosomal proteins (a fixed parameter, therefore not reported in the table), necessary for enzyme functionalities and ribosome assembly,[226,227] along with tRNA and amino acids important for translation and in vitro protein synthesis were included. Besides, the reducing agent DTBA that was recently demonstrated to be important for the iSAT system fueled by the PANOx-SP system,[152,199,205] was comprised in the mixture.

While the concentration of each component can in theory be varied continuously, the resulting experimental space would be impossible to access. We therefore decided to define the experimental space with each component varied across 4 concentration levels: low, medium-low, medium-high and high (**Table 5.1**). This yields an experimental space with a total of 4^{20} (~$1.1*10^{12}$) possible experiments, which is orders of magnitude above what is possible to be executed exhaustively. Thus, our experimental exploration focused on 8 small subsets of this design space, each defined by a collection of experiments located in the neighborhood of a different "pivot" experiment in the space (**Table 5.1**). This neighborhood is defined by all experiments obtained varying one component on the four concentration levels (**Table 5.1**), while keeping all of the other components constant at the same concentration level as in the pivot, as imposed by our robotic configuration. We started with two exploratory generations (generations 1.1 and 1.2) where the pivots were manually chosen. EDoE subsequently estimated a predictive model on all experimental results and exploited this model to select the pivot for the following generation; this process was iterated five more times, for a total of eight generations (**Figure 5.2**).

5.4.2   Improving iSAT protein yield through iterated EDoE

Over the course of seven generations of experiments we established iSAT in *E. coli* BL21 Rosetta2, implemented a novel metabolic scheme, and improved protein yield 10-fold from the base case. To initially populate our EDoE dataset, medium-high and high concentration levels of components dominate the pivots of generations 1.1 and 1.2 (**Table 5.1**). Higher concentrations were selected in part because traditional iSAT in *E. coli* MRE600 uses similar concentrations. However, such molecular component configurations resulted in low response values (**Figure 5.3**; **Figure 5.4**) in the BL21Rosetta2 system.
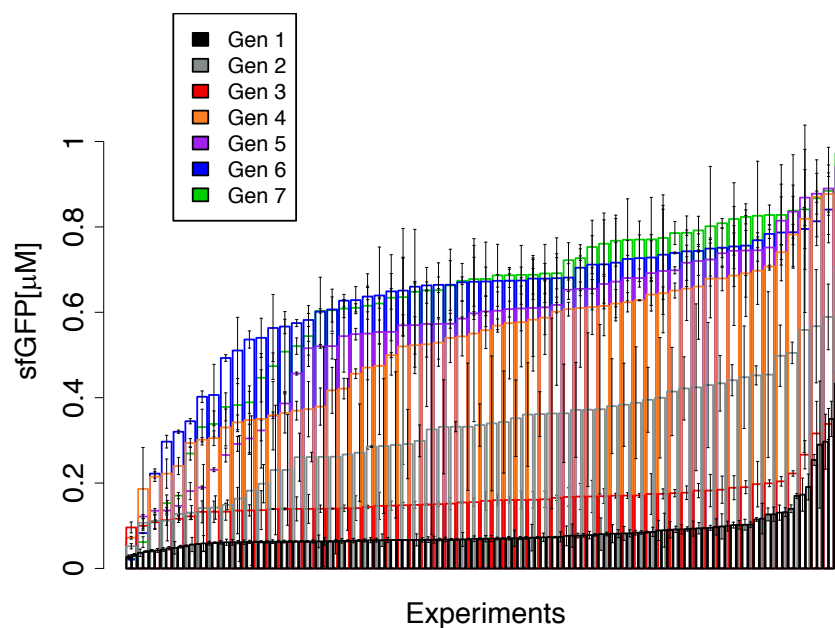
Figure 5.3. Response distribution of each generation. *Response is defined as the fluorescence signal from GFP present, converted into a µM scale. The responses of generations 1.1 and 1.2 are displayed here as a single "Gen 1" distribution. The error bars are the standard deviation of two independent samples.*
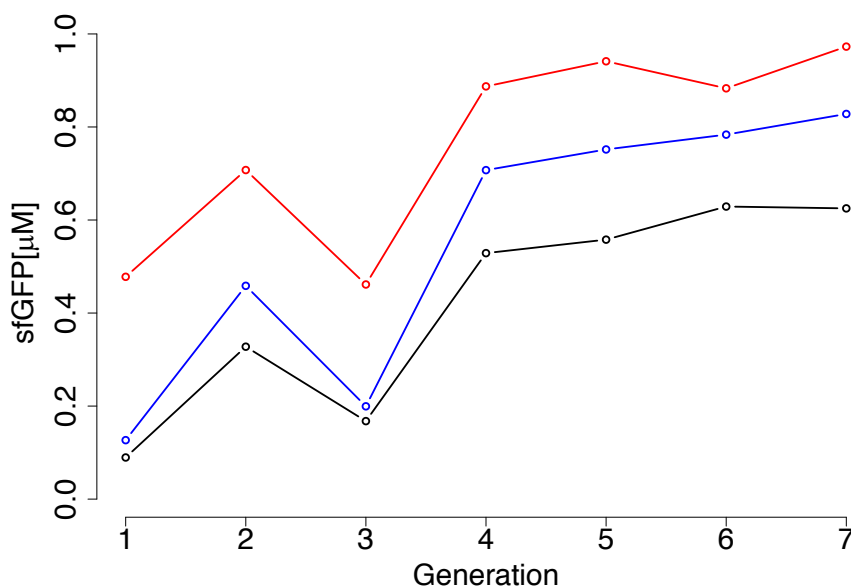
Figure 5.4. Responses by experimental generation. *The trajectories show max (red), mean (black), and 90% percentile (blue) of the response distribution in each generation of experiments. The responses of generations 1.1 and 1.2 are here lumped together in a single "generation 1".*

With the EDoE approach, we observed an overall increase in response in the subsequent generations, with the most substantial boosts obtained in generations 4–7. Indeed, the synthesized sfGFP

from our iSAT system increased from ~0.1 µM in generations 1.1-1.2 to ~1 µM in the last generation (**Figure 5.3**, **Figure 5.4**). The increased protein yields resulted from variation of the metabolic scheme, given by the concentration levels of the components in the pivots, and their neighborhoods, in different generations. The pivots showed a progressive convergence toward low or zero concentration levels in later generations. Overall, this indicates that several of the components used to design the initial pivots for high-throughput screening can be omitted, therefore providing a simplified mixture for energizing in vitro protein synthesis using in situ self-assembled ribosomes. However, two molecular components, GTP and NAD, converge to consistently high values in generation 4 and beyond. In particular, GTP is important for the translation process[228] and NAD is a cofactor in glycolysis and metabolism.[229]

5.4.3   Dependence of response on individual components

We next developed an intuition about the most crucial small molecules for iSAT in BL21Rosetta2 by investigating the impact that different concentration levels of individual components had on the experimental response, as measured in the 480 experiments that made up all generations (**Figure 5.5**). An important observation is that certain molecules appear to have a negative effect on the response at any concentration, and their omission may therefore be actually beneficial for the system. This was not intuitive a priori based on previous literature of similar systems. In particular, it seems that nucleotide monophosphates can be omitted from the reaction mixture. However, this is true to a lesser extent for UMP, which is a precursor of UTP necessary for in vitro transcription. Most likely, such precursor molecules and their triphosphate equivalents are already present in the S150 crude lysate in quantity sufficient to generate chemical energy (ATP) and initiate RNA polymerization to sustain the in vitro protein synthesis with the iSAT system in BL21Rosetta2. Conversely, it appears that GTP must be used at high concentration (8 mM) to boost the response of the reaction. Therefore, GTP can be considered as an essential high-energy molecule extremely relevant for the metabolic scheme presented herein. This molecule is important for translation and in the regeneration of nucleotides triphosphates.[230]

Phosphate donor molecules, such as acetyl-phosphate also appeared to be an important component to increase response. Indeed, higher protein yields were observed with 0.32 mM acetyl-phosphate as compared to 0 mM. Acetyl-phosphate is important as high-energy phosphate donor for ATP

and GTP-regeneration systems[205,230] and has been shown to be useful for efficient NTPs and dNTPs regeneration in E. coli lysates.[231] Conversely, increasing the concentration of the phosphate buffer K2HPO4/KH2PO4 resulted in a gradual decrease in response. Indeed, if inorganic phosphate is accumulated at high concentration in the cell-free reaction, it inhibits the cell-free reaction, mainly by sequestering magnesium that is necessary for protein biosynthesis.[232] However, a low concentration is slightly beneficial. The inorganic phosphate would trigger the hydrolysis of maltodextrin to activate the glycolytic pathway in the crude extract.[208] Interestingly, as the system does not need expensive molecules such as 3PGA and PEP to re-generate ATP, we could conclude that it is energized mainly by the ATP produced through the glycolytic pathway using the polysaccharide (maltodextrin) as the carbon source, in addition to GTP and acetyl-phosphate high-energy phosphate donor molecules.
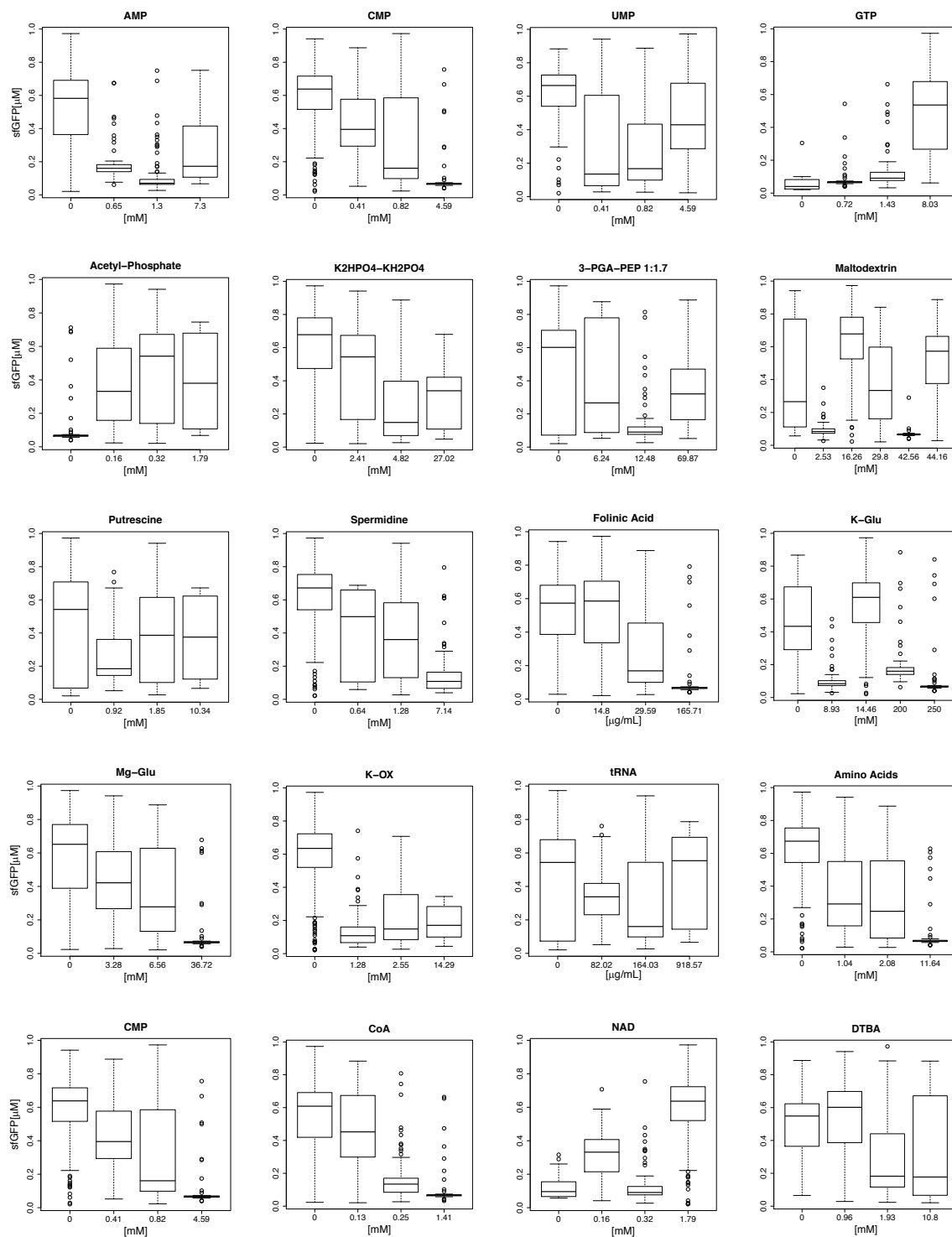
Figure 5.5. Dependence of iSAT response on individual experimental components. *Each of the 20 components tested in iSAT is plotted in a separate graph. The x-axis represents concentration levels and the y-axis the corresponding conditional response distributions. Results are aggregated over the experiments performed in all generations.*

Concerning polyamines, i.e. putrescine and spermidine, we observed a general and gradual decrease in the sfGFP synthesis at increasing concentrations. The same trend was observed for folinic acid, which is a molecular component used in the design of the reaction buffer of cell-free expression systems based on 3-PGA.[207,208] The response shows a somewhat non-linear dependence on K-glutamate, with higher response at 0 mM or at 14 mM. The dependence of the response on Mg-glutamate is instead more linear, with a decline of the response proportional to the increase in Mg-glutamate in the system. Magnesium salts are important for ribosome stability in *E. coli* cell-free expression system,[233] as well as enzymes functionalities, and their concentration must be tuned to an optimal value to avoid detrimental effects on the system.[234] It should be mentioned that putrescine, spermidine, K-glutamate, and Mg-glutamate are components already included in the system during crude extract preparation, and therefore present in the S150 cell extract (see material and methods). This could explain higher response values when such components are not in the chemical mixture making the reaction buffer for protein synthesis. Moreover, it also appears that the omission of K-oxalate from the reaction buffer can result in higher system response. Normally, this salt is important to inhibit the reverse reaction of the phosphoenolpyruvate synthase with PEP or 3-PGA used as high-energy phosphate molecule donor.[205,235] Therefore, this finding suggests that the S150 system can by-pass the substrate level phosphorylation for ATP regeneration.

An overarching design principle that emerged was that components involved directly in the translation of proteins need to be finely-tuned. For instance, the response tends to be quite sensitive to differences in the concentration of tRNA. This molecule is already present in the crude extract after preparation and already charged with its correspondent amino acid. Furthermore, the response decreases with higher amino acids concentrations. cAMP and CoA show a similar trend. However, at its highest concentration (1.79 mM), NAD boosted protein synthesis and therefore appears to be an essential component for the iSAT system. NAD is an important cofactor involved in glycolysis and in turn the maltodextrin-based metabolism to energize *in vitro* protein synthesis. The reducing agent, DTBA, has been recently demonstrated important in the optimization of the iSAT system.[198] We observed that the addition at low concentration is also beneficial for the efficiency of the system presented here.

5.4.4 Composition of iSAT reaction conditions achieving highest protein expression and their kinetic profiles

The composition of iSAT reaction conditions achieving highest protein expression is shown in **Table 5.3**. These conditions show similar features. In particular, they are characterized by: AMP, K-OX, and CoA at 0 mM (with the exception of 1 experiment with CoA at 0.13 mM); GTP at 8 mM; NAD at 1.79 mM; Acetyl-phosphate at 0.32 mM (with the exception of 2 experiments at 0.16 mM); K-Glu at 14.46 mM (with the exception of one experiment at 200 mM). The analysis of these reactions highlights the importance of GTP, Acetyl-phosphate, NAD, UMP and maltodextrin as molecular framework fueling the iSAT system prepared from *E. coli* BL21Rosetta2.

Table 5.3. Top 9 combinations mean response and composition. *The mean response values are obtained by averaging out the twenty repeats within each of the two plates and then averaging out the two within-plate means; between-plate response standard deviation, calculated over the two within-plate means.*

Pivot Generation	7	5	5	4	7	6	5	4	4
Mean Response (RFU)	2594.5	2510	2373	2366	2358.5	2355	2341	2338.5	2322
% Noise	0.11	0.03	0.06	0.14	0.11	0.16	0.05	0.06	0.01
AMP	0	0	0	0	0	0	0	0	0
UMP	4.59	0.41	0.41	0.82	4.59	0	0.41	0.82	0.82
CMP	0.82	0	0	0.41	0.82	0	0	0.41	0.41
GTP	8.03	8.03	8.03	8.03	8.03	8.03	8.03	8.03	8.03
Ac.phos	0.16	0.32	0.32	0.32	0.16	0.32	0.32	0.32	0.32
K_2HPO_4/KH_2PO_4	0	2.41	2.41	4.82	0	2.41	2.41	4.82	4.82
Maltodextrin	16.26	0	16.26	44.16	16.26	16.26	44.16	44.16	44.16
3-PGA:PEP1:1.7	0	0	0	69.87	0	0	0	6.24	0
Folinic.Acid	14.8	0	0	29.59	14.8	14.8	0	29.59	29.59
Putrescine	0	1.85	1.85	1.85	0	0	1.85	1.85	1.85
Spermidine.	0	1.28	1.28	0	0	0	1.28	1.28	1.28
K.Glu	14.46	14.46	14.46	14.46	200	14.46	14.46	14.46	14.46
K.OX	0	0	0	0	0	0	0	0	0
Mg-Glu	0	3.28	3.28	6.56	0	6.56	3.28	6.56	6.56
tRNA	0	164.03	164.03	0	0	0	164.03	0	0
Amino acids	0	1.04	1.04	2.08	0	0	1.04	2.08	2.08
cAMP	0	0.66	0.66	0.66	0	0	0.66	0.66	0.66
CoA	0	0	0	0	0	0.13	0	0	0
NAD	1.79	1.79	1.79	1.79	1.79	1.79	1.79	1.79	1.79
DTBA	1.93	0.96	0.96	0	1.93	10.8	10.8	0	0

In addition, we performed a kinetic analysis to observe potential, subtle differences between our top 9 experiments (**Figure 5.6**). From the protein time-courses, we observed a window between 5 and 10

hours where iSAT reactions show different rates. Most likely, these rates depend on the ability to regenerate chemical energy to sustain RNA transcription, ribosome self-assembly, and translation. During this time most of the cell-free reactions show a lag-phase, which was not observed using the traditional iSAT system design.[198-200] Potentially, the maltodextrin-based metabolic scheme and slow energy release could also cause the difference in rate from traditional iSAT systems prepared with MRE600.
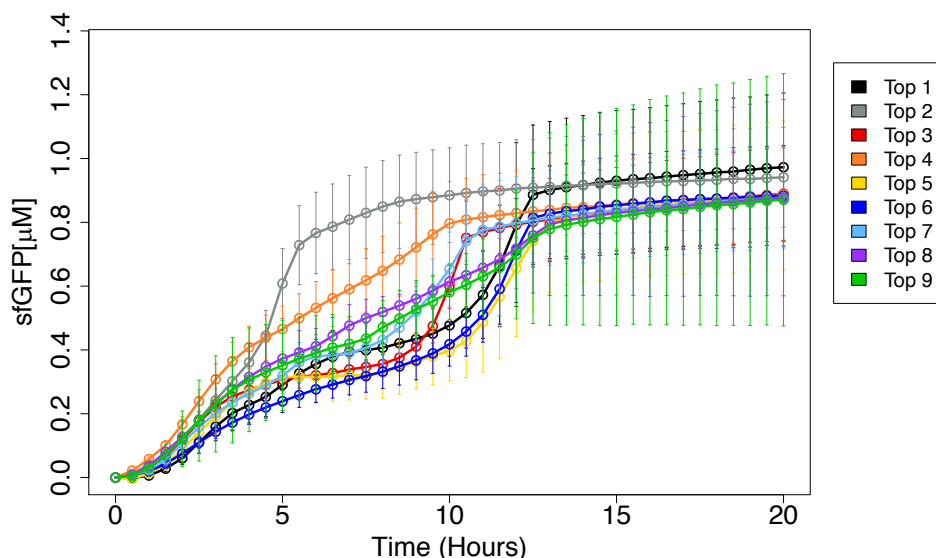
Figure 5.6. Kinetics of the top 9 iSAT experiments. *Measurements of sfGFP reporter protein (µM) collected at 30-minute intervals for each of the top 9 experiments found with EDoE. The error bars are the standard deviation of two independent samples.*

5.4.5   Variation of magnesium and DNA templates concentrations

A magnesium optimization of S150 extracts for reporter protein synthesis activity in iSAT reactions is vital for in vitro transcription/translation.[200] Therefore, following the computationally-guided EDoE optimizations, we decided to carry out an additional magnesium optimization to improve activity. We focused on optimizing buffer conditions that resulted in the top 4 protein synthesis yields (**Figure 5.6**). In addition to magnesium, we explored the effects of changing DNA concentration, since DNA concentrations used in our high-throughput experimentation was 2nM and previous efforts showed that increasing DNA concentration might yield enhanced iSAT activity.[198-200] For each of the top 4 experiments, we explored the effects of modifying the magnesium glutamate and plasmid DNA concentrations to see if iSAT activity could be improved in the context of the new metabolic scheme (**Figure 5.7**). To optimize these two parameters,

a lattice experimental design was carried out. Based on the need for greater rRNA transcription, pT7rrnB (plasmid DNA) concentration was varied with 2, 4, 8 and 10 nM of DNA. Magnesium glutamate concentration was altered by testing a range of concentrations: 0, 2, 3.3, 4, 5, 6, 6.6, 8 and 9 mM. Standard 15 µL batch reactions were prepared by hand and performed at 37 °C for 20 hours, with varying magnesium glutamate concentrations for each reaction. Collectively, these experiments allowed us to map component landscapes to explore global and local optima. Global optima for magnesium glutamate and DNA concentrations for each iSAT conditions were determined. The optimum concentration of DNA was as follows: for top experiment 1 the optimum was 2 nM, for top experiment 2 the optimum was 2 nM, for top experiment 3 the optimum was 8-10 nM and for top experiment 4 the optimum was 2 nM. For all experiments, the original magnesium glutamate concentration from each of the top 4 experiments was used. Taken together, our results show that the cell-free framework and its barrier-free access to reaction conditions is well-suited for rapidly acquiring physiochemical landscapes to assess and optimize pathway performance. This joins an emerging body of literature highlighting the value of cell-free systems for prototyping biological systems.[236,237]
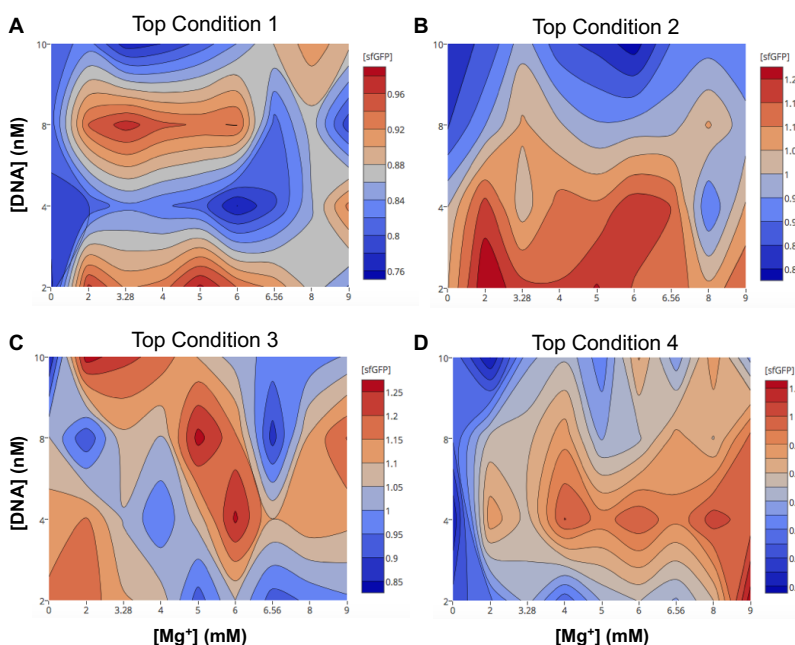
Figure 5.7. DNA and magnesium optimizations of top four experiments. *iSAT reactions to optimize pT7rrnB DNA concentration and magnesium concentration were run individually by hand under the small molecule conditions of the top four robot results: (A) condition 1, (B) condition 2, (C) condition 3, and (D) condition 4. To optimize these two parameters, a lattice experimental design was carried out. pT7rrnB concentration was varied with 2, 4, 8 and 10 nM of DNA. Standard 15 μL batch reactions were performed at 37 °C for 20 hours, with varying magnesium glutamate concentrations for each reaction. The response for sfGFP production is plotted in heat maps.*

5.4.6   High-dimensional inter-component synergy:

With all of our experimental data at hand, we set out to understand components having the largest impact of iSAT activity. This is important because it teaches us which components might synergistically work together to enable high activity. **Figure 5.8** tackles this analysis from a regression modeling perspective by showing the structure of a conditional inference tree, trained on all experimental data collected through generation 7 of the data. The tree, estimated via the ctree function in the R package partykit, represents a partition of the experimental space into regions with homogeneous experimental response. Each path from the root node to a given leaf node represents a different sequence of recursive bisections of the experimental space. Each bisection is defined by a specific experimental component (shown in ovals), selected to maximize the statistical association between a candidate component and the experimental response. The specific concentration of this component to maximize this association is shown on the two outgoing arcs. This concentration is selected to maximize the standardized difference between

the response means in the two resulting data subsamples. Each leaf node represents a different region of

the experimental space (shown in boxes) and contains a box plot of the response distribution of the

experiments belonging to that region. From this tree analysis, we find that the highest-response

experiments are, on average, mainly contained in nodes 15 and 16. This region is defined by the presence

of five components: GTP, NAD, K-Ox, 3-PGA-PEP, and Maltodextrin. Thus, these components represent

"key components" with "key concentrations" in specific intervals to improve iSAT performance.
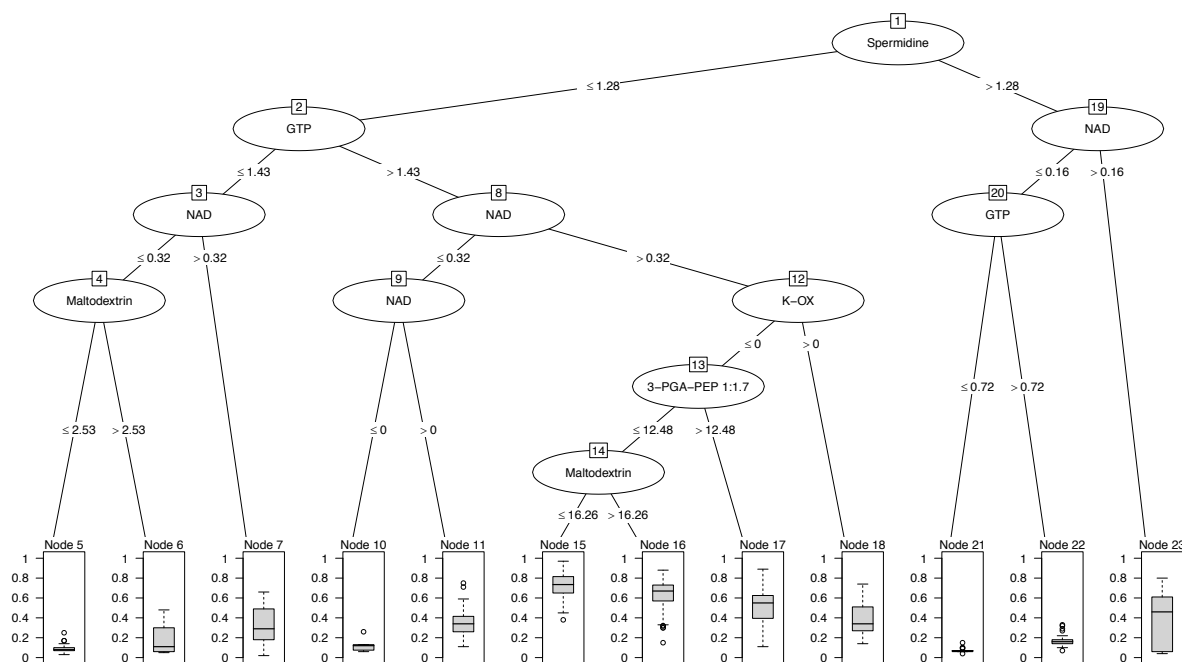
Figure 5.8. Conditional inference regression tree trained on all experimental data. Terminal *nodes show the box plots of the experimental response, conditioned on the experimental region defined by specific experimental components at specific concentrations; regions containing high-response experiments are defined by at least 4/5 experimental components.*

Further insight into the impact of these key components and key concentrations on experimental

response can be found in **Figure 5.9**. This figure shows experimental response distributions conditioned

on the number of key components at key concentrations in an experiment and illustrates how a substantial

boost in experimental response is obtained only when several of these conditions are satisfied

simultaneously. The average response, for example, fluctuates around 0.10 if two or less conditions are

satisfied, but progressively grows to 0.30, 0.49, and 0.68 as the number of satisfied conditions increases

to three, four, and five. This indicates that these key components are synergistic, that is, they yield high

response only by working together. Identifying which components are key for iSAT performance allows us to define these components as essential when trying to minimize which components we include in iSAT to minimize the cost of a reaction. Indeed, with this knowledge in hand we were able to decrease the cost of this new iSAT reaction 4-fold from the state-of-the-art reaction (**Figure 5.10**).
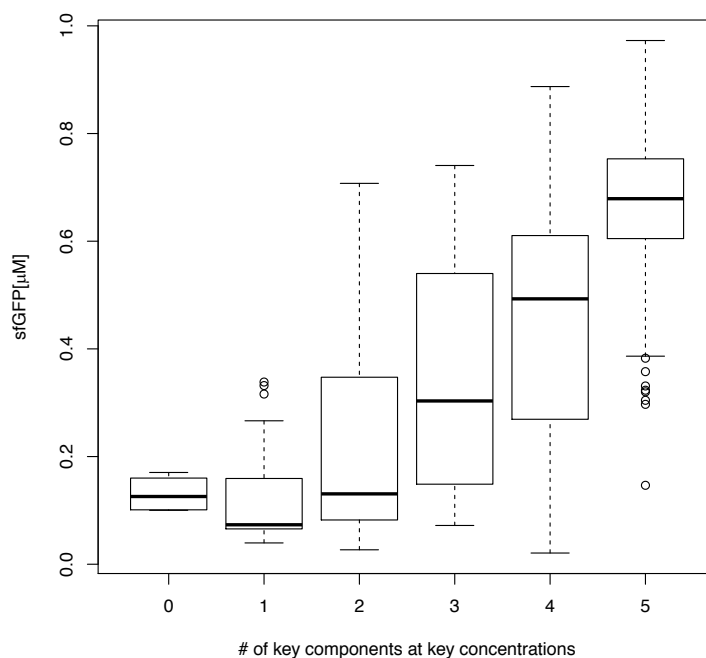
Figure 5.9. Distributions of experimental response. *Experimental response distributions conditioned on the number of experimental coordinates matching those of the high-response experiments in nodes 15 and 16 of the tree in* **Figure 5.3**. *The substantial shift of the rightmost distribution with respect to the others suggests the presence of 5-dimensional synergies in the experimental response landscape.*
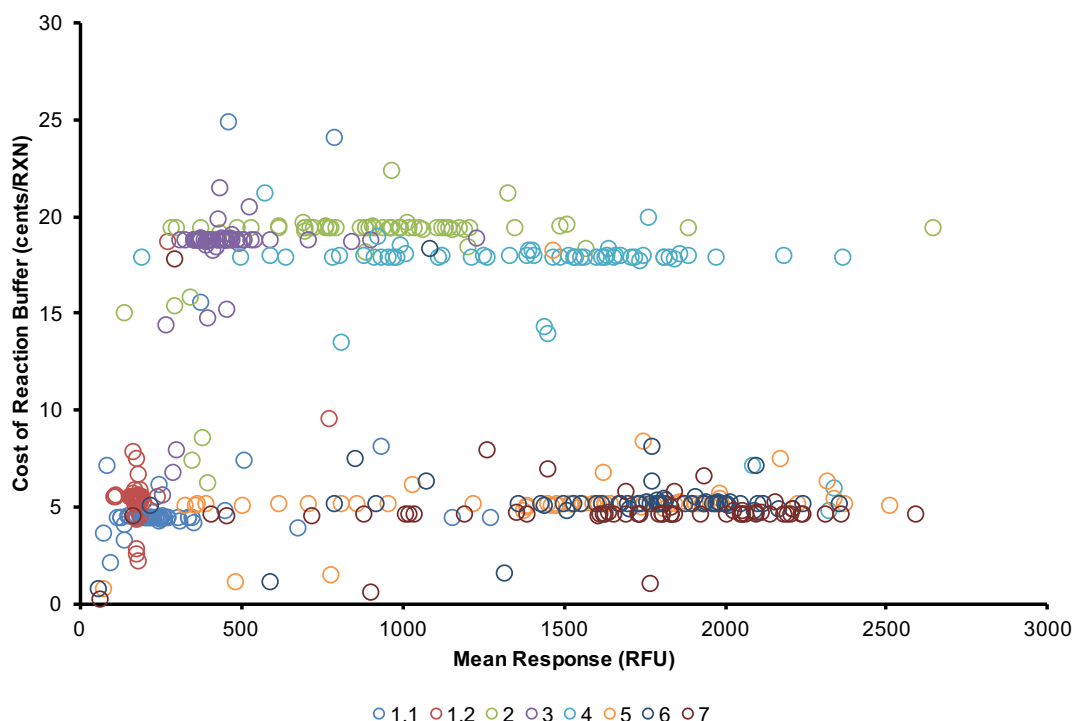
Figure 5.10. Cost of iSAT Reactions. *All conditions tested in this work were used to calculate the cost per reaction for each. The cost was determined by adding the concentration of cost of each small molecule's concentration in each condition. The cost of each small molecule was calculated from the cost of reagent from SigmaAldrich with the concentration used in each condition accounted for. Each circle in the graph represents a unique condition.*

5.5 Discussion

In this work, we have *(i)* established a novel *E. coli* bacterial strain, BL21Rosetta2 as novel platform for the iSAT system, which allows *in vitro* study of ribosome assembly and function of a new bacterial strain, *(ii)* integrated a novel simplified metabolic scheme for iSAT that exploits maltodextrin as non-phosphorylated energy source, and (iii) implemented nucleotides triphosphate regeneration for *in vitro* transcription. This is an important point as it shows a metabolic scheme with energy regeneration, transcription, translation, ribosome construction, and components necessary for self-maintenance[189,202,238]—a system that could be integrated as a sub-system of a minimal cell. Herein, the challenges were (i) the activation of such a complex metabolic scheme using the iSAT system, which more than using self-assembled ribosomes, is also diluted three times more than conventional cell-free protein

synthesis systems. Therefore, these results could be also important to scale-up cell free ribosome synthesis for medical and industrial applications,[161] such as the synthesis of peptidomimetics.

In addition, we have demonstrated that our EDoE approach can be adapted to the technical constraints of a robotic workstation, which in this work substantially limited the exploration of the experimental components, unlike in previous applications of EDoE.[211,212] Our EDoE algorithm was able to optimize the iSAT system into an *E. coli* strain sub-optimal for studying ribosome assembly, discovering complex inter-component synergies, while also decreasing the cost of the cell-free reaction 4-fold, i.e. from ~20 cents to 5 cents per reaction (**Figure 5.10**).

Reaction optimization may be specific to which target proteins are made in the reactions. Because we used a single target protein, sfGFP, to test the yield of the iSAT protein synthesis protocol, the concentrations selected by the EDoE algorithm, albeit optimal for this particular protein, may be sub-optimal for other proteins. This is a form of "overfitting", and can only be addressed with further experiments, using other target proteins. In addition, we believe that further optimization of the system should comprise the design of the DNA templates, either in the length and sequence of the regulatory parts and spacer sequences.[212,225,239] Looking forward, we anticipate that new cost-effective iSAT reactions fueled by new energy regeneration schemes discovered here, could facilitate unraveling the systems biology of ribosome biogenesis, constructing minimal cells from defined components, and engineering ribosomes with new functions

5.6 Conclusions

This chapter describes a series of optimizations of iSAT systems that captures my development of liquid-handling robotics for cell-free systems. While we only achieve a minimal amount of protein production capabilities after seven generations of machine-learning, we limited ourselves to a given set of components and concentrations. When you are working with such complex systems it is important to limit the scope of alterations and optimizations. Further work could and should be pursued in optimizing such cell-free systems in order for them to be more widely adopted.

Chapter 6: Controlling cell-free metabolism through physiochemical perturbations

I get obsessed by little nerdy things in my corner that no one else is interested in.

- Björk

This chapter is a transition from an initial system presented in Chapter 3 to robust system presented in Chapter 7. In developing the system used in Chapter 7, I ran into inconsistencies in how well a biosynthetic pathway performed depending on which small molecules were added to the cell-free reactions. Rather than marching forward to bigger and seemingly more interesting research, I decided to not give up in figuring out the 'why' of these inconsistencies. At this same time, I had just taken on the mentoring of two undergraduate students. It is with their help that this chapter was possible. This story was published in Metabolic Engineering with both of them as co-authors.

6.1 Abstract

Building biosynthetic pathways and engineering metabolic reactions in cells can be time-consuming due to complexities in cellular metabolism. These complexities often convolute the combinatorial testing of biosynthetic pathway designs needed to define an optimal biosynthetic system. To simplify the optimization of biosynthetic systems, we recently reported a new cell-free framework for pathway construction and testing. In this framework, multiple crude-cell extracts are selectively enriched with individual pathway enzymes, which are then mixed to construct full biosynthetic pathways on the time scale of a day. This rapid approach to building pathways aids in the study of metabolic pathway performance by providing a

unique freedom of design to modify and control biological systems for both fundamental and applied biotechnology. The goal of this work was to demonstrate the ability to probe biosynthetic pathway performance in our cell-free framework by perturbing physiochemical conditions, using *n*-butanol synthesis as a model. We carried out three unique case studies. First, we demonstrated the power of our cell-free approach to maximize biosynthesis yields by mapping physiochemical landscapes using a robotic liquid-handler. This allowed us to determine that NAD and CoA are the most important factors that govern cell-free *n*-butanol metabolism. Second, we compared metabolic profile differences between two different approaches for building pathways from enriched lysates, heterologous expression and cell-free protein synthesis. We discover that phosphate from PEP utilization, along with other physiochemical reagents, during cell-free protein synthesis-coupled, crude-lysate metabolic systems inhibits optimal cell-free butanol metabolism. Third, we show that non-phosphorylated secondary energy substrates can be used to fuel cell-free protein synthesis and *n*-butanol biosynthesis. Taken together, our work highlights the ease of using cell-free systems to explore physiochemical perturbations and suggests the need for a more controllable, multi-step, separated cell-free framework for future pathway prototyping and enzyme discovery efforts.

6.2   Introduction

Building and optimizing biosynthetic pathways in cells offers promising solutions to problems in energy and medicine but is often time-consuming.[5,10,163] This is, in part, because pathway construction involves designing and tuning DNA cassettes to balance heterologous enzyme expression levels, selecting the best enzymes for the desired chemistries, and engineering native metabolic pathways to channel flux towards the desired product.[163,240] In addition, enzyme candidates from databases (like BRENDA,[241] METACYC,[242] etc.) containing thousands of enzyme variants with kinetic data available must be selected and tested. However, with even the most informed selection process, it may be that the selected enzymes do not have the desired properties for a particular biosynthesis.[141,243] With the current metabolic engineering Design-Build-Test (DBT) cycle time being on the order of weeks to months to test a given pathway configuration, it can take hundreds of person-years of research and development time to test the number of enzyme and pathway variants needed to bring a biochemical to market.[13,143,240]

Cell-free systems provide many advantages for accelerating DBT cycles and probing metabolism.[138,144,244] For example, the open reaction environment allows direct monitoring and manipulation of the system to study pathway performance. As a result, many groups have used purified enzyme systems to study enzyme kinetics and inform cellular expression.[44,46,144] Recently, the Liao group showed that cell-free pathways could be used as a tool to successfully develop a non-oxidative glycolytic pathway that maximizes carbon utilization in *Escherichia coli* whole cells.[46] Despite advances in purified enzyme systems, crude lysates are becoming an increasingly popular alternative to build biosynthetic pathways because they inherently provide the context of native-like metabolic networks[72,171-173].

The presence of native metabolic enzymes and cofactor regeneration in crude lysates[152] allows for observations of metabolic interactions with biosynthetic pathways, which is limited in purified systems.[144] For instance, the Panke group has shown that DHAP can be made in crude lysates and real-time monitoring can optimize production.[72,172] In addition, our group has shown that 2,3-butanediol,[121] mevalonate,[171] *n*-butanol,[173] and more complex products[174] can be constructed in crude lysates with high productivities (>g/L/h). In these systems, removing cell growth requirements reduces the complexity of the metabolic reactions that could interact with the expression of a biosynthetic pathway while maintaining native energy metabolism active. Crude lysates also have the capability to construct biosynthetic pathways by expressing enzymes directly *in vitro* by cell-free protein synthesis (CFPS) (*in vitro* transcription and translation).[173,174]

Producing chemicals and testing biosynthetic pathways via crude lysate metabolic engineering is quick and efficient. In our recent work[173] with the CoA-dependent *n*-butanol pathway (**Figure 6.1A**) we developed a cell-free framework to study metabolic pathways. In our framework, cell-free cocktails for synthesizing target small molecules are assembled in a mix-and-match fashion from selectively enriched crude-cell lysates containing pathway enzymes either from heterologous overexpression prior to lysis (**Figure 6.1B**) or from direct *in vitro* production by CFPS (**Figure 6.1C**). Incorporating CFPS as a means of enzyme production reduces the time to build pathways and test heterologous enzymes. However, coupling CFPS with crude extract mixing for building biosynthetic pathways (CFPS-ME) underperforms in the context of *n*-butanol production titers, yields and productivities as compared to the mix-and-match pre-enriched lysates (CFME) approach.
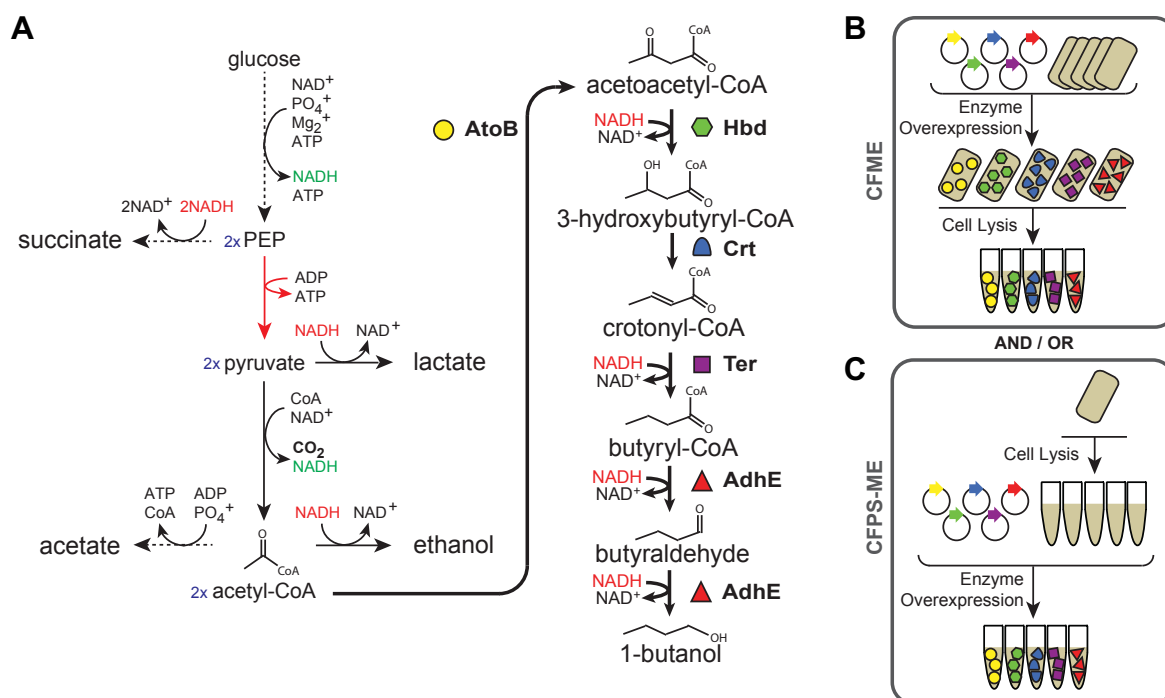
Figure 6.1. A schematic for cell-free expression of a n-butanol model pathway. *(A) Schematic representation of the constructed biosynthetic* n-*butanol pathway. Acetyl-CoA is generated through* E. coli*'s natural glycolysis and funneled into the* C. acetobutylicum-*derived CoA-dependent pathway to produce* n-*butanol. Major by-products and cofactors are shown. Heterologous enzymes used are as follows: AtoB (*Escherichia coli*), Hbd (*Clostridia beijerinckii, CB*), Crt (*Clostridium acetobutylicum, CA*), Ter (*Treponema denticola, TD*), AdhE (*Clostridium acetobutylicum, CA*). Methodologies for enzyme production for (B) cell-free metabolic engineering (CFME) and (B) cell-free protein synthesis driven metabolic engineering (CFPS-ME) are depicted.*

It would be beneficial if we could probe how native metabolism and heterologous pathways interact in crude lysates through physiochemical alterations both to understand why CFPS-ME underperforms and to increase the utility of the cell-free platform to study cellular metabolism and biosynthesis. We would be able to study how enzymes might synergize with each other and identify ratios of enzymes[171] necessary for optimal pathway performance. We could learn which regimes of metabolism in cell-free might mimic or relate to cellular environments. Additionally, we could probe enzyme inhibition mechanisms by manipulation of the physiochemical environments.

In this work, we set out to demonstrate the ability to probe biosynthetic pathway performance through physiochemical parameters, using cell-free *n*-butanol synthesis as a model. The key idea was to show that we could rapidly explore and influence cell-free pathway performance by varying unique combinations of substrates, cofactors, salts, buffers, etc., in crude extract-based cell-free

biotransformations. We specifically carried out three distinct case studies focused on how by-products evolve over the time of a reaction. First, we wanted to demonstrate that such studies could be enabled using robotic liquid-handling systems to access combinatorial design space that exceeds typical pipelines pursued in cells. By testing 40 different reaction conditions in one day, we identified NAD and CoA as the two most important parameters controlling cell-free *n*-butanol biosynthesis and improved titers 9.3 ± 2.3-fold as compared to our starting conditions. We then generated cofactor landscape maps that identify optimum cofactor ratios for pathway performance. Second, we wanted to rigorously study why and how we might mitigate the underperformance of CFPS-coupled crude-lysate metabolic systems to better understand how these systems can be used to rapidly prototype biosynthetic pathways. Third, we wanted to show that different energy sources could be used to fuel CFPS and cell-free *n*-butanol production to relieve phosphate toxicity. Taken together, our results show that cell-free systems enable combinatorial and modular assembly of pathways to improve pathway performance using well-defined experimental conditions. They also set the stage for the development of a multi-step, separated CFPS-coupled cell-free metabolic engineering framework for controlled enzyme and biosynthetic pathway study.

6.3 Materials and Methods

6.3.1 Bacterial strains and plasmids

Both *E. coli* DH5α and NEB Turbo™ (NEB) were used for plasmid preparation. *E. coli* BL21(DE3) (NEB) was used for protein overexpression and for preparation of all extracts. A modified version of pET-22b (Novagen/EMD Millipore), used in previous studies,[121,173] was used for all constructs for *in vivo* overexpression of proteins. For *in vitro* expression of proteins, the pJL1 vector was used. Carbenicillin (100 µg ml^{-1}) was used with the pET vector system and kanamycin (50 µg ml^{-1}) was used with the pJL1 vector system. Propagated constructs were purified using an EZNA Plasmid Mini Kit (Omega Bio-Tek). All plasmids used in this study were used in our previous studies.[173]

6.3.2   Cell Extract Preparation

$E.\ coli$ BL21(DE3) cells were grown in 2 × YTPG media (16 g l^{-1} tryptone, 10 g l^{-1} yeast extract, 5 g l^{-1} NaCl, 7 g l^{-1} potassium phosphate monobasic, 3 g l^{-1} potassium phosphate dibasic, 18 g l^{-1} glucose). These cells were cultured at the 50 ml scale in 250 ml baffled tunair shake flasks (IBI Scientific, Peosta, IA) and at the 1 L scale in a 37 °C incubator with shaking at 250 rpm. When cells reached OD_{600} = 0.6-0.8, the cultures were induced with 0.1 mM IPTG. After induction cultures were grown for 4 h at 30 ºC when plasmid expression was required and grown to OD_{600} = 3.0 for plasmid-less extracts. Antibiotics were not used during cell growth. The cells were harvested by centrifugation at 8,000g at 4 °C for 15 min and were washed two times with cold S30 buffer (10 mM Tris-acetate (pH 8.2), 14 mM magnesium acetate, and 60 mM potassium glutamate). After final wash and centrifugation, the pelleted wet cells were weighed, flash frozen in liquid nitrogen, and stored at −80°C. The thawed cells were suspended in 0.8 ml of S30 buffer per 1 g of wet cell mass. In order to lyse cells by sonication, thawed and suspended cells were transferred into 1.5 ml microtube and placed in an ice-water bath to minimize heat damage during sonication. The cells were lysed using a Q125 Sonicator (Qsonica, Newtown, CT) with 3.175 mm diameter probe at frequency of 20 kHz and 50% of amplitude. The input energy (Joules) was monitored and 830 J was used for 1.4 ml of suspended cells.[169] The lysate was then centrifuged at 12,000g at 4 °C for 10 min. All of prepared cell extracts were flash frozen in liquid nitrogen and stored at −80 °C until use. Quick-Start Bradford protein assay kits (Bio-Rad) were used to measure the total protein concentration of each extract with a bovine serum albumin standard.

6.3.3   CFME Reactions

Reactions were carried out in 1.5 ml Eppendorf tubes at 37 °C in 25 µl volumes. Each reaction consisted of mixing five extracts, containing one enzyme overexpressed each, to complete the biosynthetic n-butanol pathway (2 mg ml^{-1}) along with magnesium glutamate (8 mM), ammonium glutamate (10 mM), potassium glutamate (134 mM), glucose (120 mM), dipotassium phosphate (10 mM, pH 7.2), Bis Tris (100 mM), NAD (3 mM), and CoA (1.5 mM), unless otherwise noted. Reactions were terminated by adding 5% w/v trichloroacetic acid in a 1:1 ratio. Precipitated proteins were pelleted by centrifugation at 21,000g for 10 min at 4°C. The supernatant was stored at −80 °C until analysis.

6.3.4   Liquid-Handling Robotics

An Agilent BRAVO liquid-handling robot (Agilent Technologies) was used to carry out the reaction setup for select CFME reactions. The liquid-handling workstation has nine plate decks and a 96-pipette tip head movable in the x-y-z directions. The workstation was programmed using VWorks™ Automation Control Software (BioNex Solutions, Inc.) to pipette different arrangements of reagents. Each CFME reagent (magnesium glutamate, ammonium glutamate, potassium glutamate, glucose, dipotassium phosphate, NAD, ATP, and CoA) was diluted to five different concentrations using dH_2O and stock solutions in a 96-well plate. Each component was then pooled together into one well on a new plate so that each well in the new plate contained all eight components at standard concentrations except one reagent that was varied at a time (**Figure 6.2A**). Therefore, the new plate contained 40 wells with one unique reagent and concentration tested in each. CFME mixed extracts containing the full *n*-butanol pathway were then added to each well. All reaction conditions set up are listed in **Table 6.1**.

Table 6.1. Physiochemical conditions assembled by the liquid-handling robot.

Condition	Reagent Added (mM)							
#	Mg(Glu)$_2$	NH$_4$(Glu)	K(Glu)	Glucose	K$_2$HPO$_4$	NAD	ATP	CoA
1	8.0	10.0	134.0	0.0	10.0	0.5	0.0	0.5
2	8.0	10.0	134.0	200.0	10.0	0.0	0.0	0.5
3	8.0	10.0	134.0	200.0	10.0	0.5	0.0	2.0
4	8.0	10.0	134.0	200.0	10.0	0.5	0.0	1.6
5	8.0	10.0	134.0	200.0	10.0	0.5	2.0	0.5
6	8.0	10.0	134.0	200.0	20.0	0.5	0.0	0.5
7	8.0	10.0	134.0	200.0	10.0	0.5	0.0	1.2
8	8.0	10.0	0.0	200.0	10.0	0.5	0.0	0.5
9 (base case)	8.0	10.0	134.0	200.0	10.0	0.5	0.0	0.5
10	8.0	10.0	134.0	200.0	10.0	0.5	0.0	0.8
11	8.0	10.0	35.0	200.0	10.0	0.5	0.0	0.5
12	8.0	10.0	134.0	200.0	10.0	0.5	1.2	0.5
13	8.0	10.0	134.0	200.0	10.0	0.4	0.0	0.5
14	8.0	10.0	134.0	200.0	16.0	0.5	0.0	0.5
15	8.0	10.0	175.0	200.0	10.0	0.5	0.0	0.5
16	8.0	10.0	134.0	200.0	10.0	0.5	0.0	0.5
17	8.0	10.0	134.0	200.0	10.0	0.5	1.6	0.5
18	16.0	10.0	134.0	200.0	10.0	0.5	0.0	0.5
19	8.0	10.0	134.0	120.0	10.0	0.5	0.0	0.5
20	3.2	10.0	134.0	200.0	10.0	0.5	0.0	0.5
21	8.0	10.0	105.0	200.0	10.0	0.5	0.0	0.5
22	8.0	10.0	134.0	200.0	12.0	0.5	0.0	0.5
23	8.0	10.0	134.0	200.0	10.0	0.5	0.0	0.0
24	12.8	10.0	134.0	200.0	10.0	0.5	0.0	0.5
25	8.0	10.0	140.0	200.0	10.0	0.5	0.0	0.5
26	8.0	10.0	134.0	200.0	10.0	0.5	0.0	0.4
27	8.0	10.0	134.0	40.0	10.0	0.5	0.0	0.5
28	8.0	3.2	134.0	200.0	10.0	0.5	0.0	0.5
29	8.0	10.0	134.0	160.0	10.0	0.5	0.0	0.5
30	8.0	0.0	134.0	200.0	10.0	0.5	0.0	0.5
31	9.6	10.0	134.0	200.0	10.0	0.5	0.0	0.5
32	8.0	12.8	134.0	200.0	10.0	0.5	0.0	0.5
33	8.0	10.0	134.0	200.0	10.0	0.8	0.0	0.5
34	8.0	16.0	134.0	200.0	10.0	0.5	0.0	0.5
35	8.0	10.0	134.0	200.0	0.0	0.5	0.0	0.5
36	8.0	10.0	134.0	200.0	4.0	0.5	0.0	0.5
37	8.0	9.6	134.0	200.0	10.0	0.5	0.0	0.5
38	8.0	10.0	134.0	200.0	10.0	2.0	0.0	0.5
39	8.0	10.0	134.0	200.0	10.0	1.2	0.0	0.5
40	8.0	10.0	134.0	200.0	10.0	1.6	0.0	0.5

6.3.5   CFPS-ME Reactions

CFPS reactions were performed to express enzymes involved in *n*-butanol production prior to starting the CFME portion of the reactions using a modified PANOx-SP system.[152,153] A 25 µl CFPS reaction in a 1.5 ml microcentrifuge tube was prepared by mixing the following components: ATP (1.2 mM); GTP, UTP, and CTP (0.85 mM each); folinic acid (34.0 µg ml^{-1}); *E. coli* tRNA mixture (170.0 µg ml^{-1}); 20 standard amino acids (2 mM each); nicotinamide adenine dinucleotide (NAD; 0.33 mm); coenzyme-A (0.27 mM); spermidine (1.5 mM); putrescine (1 mM); potassium glutamate (130 mM); ammonium glutamate (10 mM); magnesium glutamate (12 mM); phosphoenolpyruvate (PEP; 33 mM), and cell extract (~12 mg ml^{-1}). For each reaction plasmid was added at ~13.3 or ~26.6 µg ml^{-1}. Reactions were incubated at 30 °C for 3 h. The *n*-butanol production portion of the reaction was initiated by supplementing glucose (120 mM) and additional reagents (3 mM NAD and 1.5 mM CoA final concentrations) noted throughout the manuscript and were incubated at 30 °C for 24 h. The final concentration of cofactors in Figure 3 CFPS-ME reactions was 0.3 mM NAD and 1.5 mM CoA, matching CFPS-ME reaction conditions previously published.[173] The total extract concentration during the *n*-butanol production portion of the reaction is 10 mg ml^{-1}, the same as a CFME reaction.

6.3.6   GFP Quantification

Superfolder GFP (sfGFP) is a reporter protein used during CFPS from the plasmid pJL1-sfGFP. Active sfGFP protein yields were quantified by measuring fluorescence. Two microliters of CFPS reaction was added in the middle of the flat bottom of 96-well half area black plates (Costar 3694; Corning Incorporated, Corning, NY). sfGFP was excited at 485 nm while measuring emission at 528 nm with a 510 nm cutoff filter. The fluorescence of sfGFP was converted to concentration (µg ml^{-1}) according to a standard curve.[32]

6.3.7   Quantification of Metabolites

High-performance liquid chromatography (HPLC) was used to analyze glucose, lactate, acetate, and *n*-butanol in the reactions. The metabolites were measured with an Agilent 1260 series HPLC system (Agilent, Santa Clara, CA) via a refractive index (RI) detector. Analytes were separated using an Aminex

HPX-87H anion exchange column (Bio-Rad Laboratories) with a 5 mM sulfuric acid mobile phase at 55 °C and a flow rate of 0.6 ml min^{-1}. Commercial standard of each metabolite was used for quantification of experimental samples by linear interpolation of external standard curves.

6.4 Results

The goal of this work was to demonstrate the ability to rapidly explore and influence cell-free metabolism and *n*-butanol synthesis by varying unique combinations of physiochemical parameters in three ways. First, we tested the effect of each small molecule reagent added to the cell-free system on cell-free *n*-butanol production with enzymes pre-enriched by heterologous expression (CFME) and optimized the small molecule conditions to increase *n*-butanol production. Second, we compared the metabolic profiles of our CFME system with that of the CFPS-coupled system (CFPS-ME) and tested the individual effect of each reagent used during CFPS on *n*-butanol synthesis. Third, we attempted to mitigate negative effects of CFPS by using non-phosphorylated secondary energy substrates to fuel cell-free protein synthesis and *n*-butanol biosynthesis.

6.4.1 Changes in NAD and CoA greatly contribute to changes in metabolism

We first carried out a series of optimization experiments to assess the impact of small molecule solutes on CFME *n*-butanol biosynthesis activity. The CFME reaction contains five mixed lysates pre-enriched with each individual *n*-butanol pathway enzyme (**Figure 6.1B**) along with substrates, cofactors, and salts designed to mimic the cytoplasmic environment. Specifically, we explored the effects salt concentration (*i.e.*, magnesium glutamate, potassium glutamate, and ammonium glutamate), glucose concentration, cofactor concentration (*e.g.,* NAD, CoA), phosphate concentration, and ATP concentration on *n*-butanol synthesis. While we know the concentrations of each of these components in *E. coli* cells (**Table 6.2**), varying the concentration of these components could provide insight into pathway operation. We used a robotic liquid-handling system to vary the concentrations of these eight chemical reagents (**Figure 6.2A**). The robot first makes serial dilutions of each of the eight components. Then, the robot pools

the reagent dilutions into 40 unique physiochemical environments (**Table 6.1**). To initiate *n*-butanol synthesis, the mixed extract is then added to each physiochemical condition. Reactions are incubated at 30 ºC in 96-well plates. **Figure 6.2B** shows the amount of *n*-butanol produced at 24 h in the 40 unique physiochemical conditions tested. The variability in technical replicates is due to accuracy of pipetting by the liquid-handler. In a single pass, we were able to increase *n*-butanol titers 9.3 ± 2.3-fold as compared to our originally reported CFME system conditions (black bar). This result highlights the potential power of cell-free systems to rapidly optimize pathway performance.

Table 6.2. Physiochemical conditions in a living E. coli cell.

Physiochemical Component	E. coli Cytoplasm
Magnesium (mM)	~25 (~3 free)[51]
Potassium (mM)	~ 200-1000[51]
Ammonium (mM)	~ 5-10[51]
Glutamate (mM)	~ 50-350[51]
Glucose	growth medium
Phosphate (mM)	~5[51]
NAD (mM)	2.6[245]
ATP (mM)	9.6[245]
CoA (mM)	1.4[245]
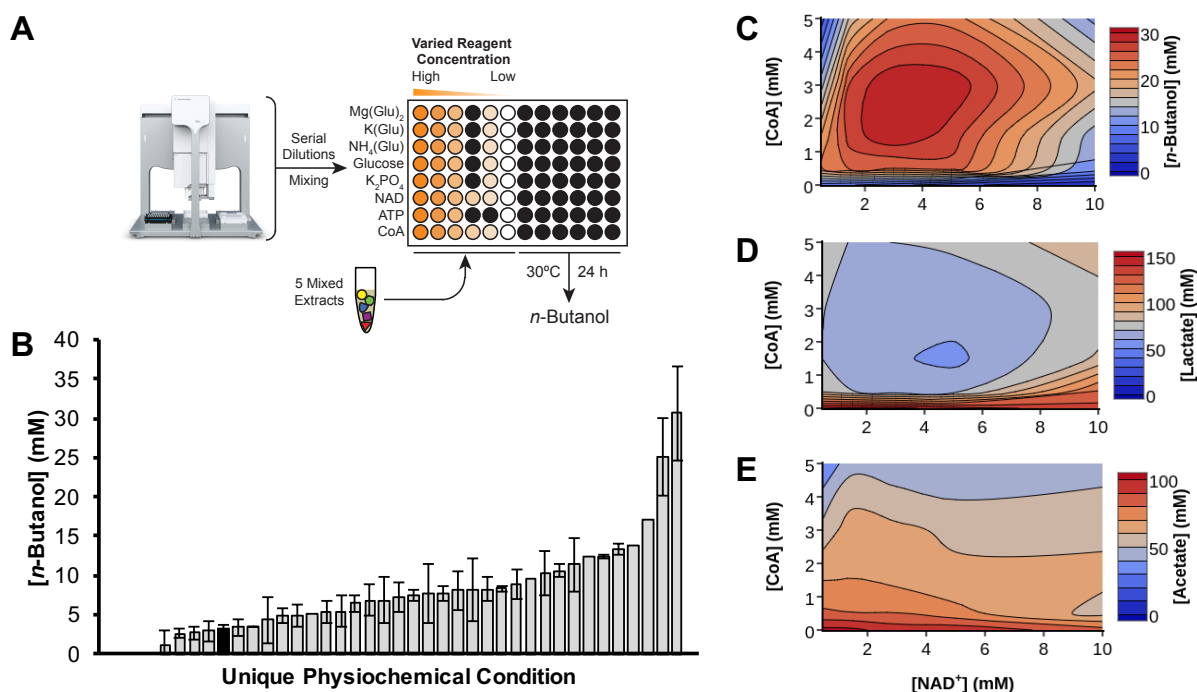
Figure 6.2. Liquid-handling robotics guide physiochemical optimizations of cell-free n-butanol metabolism. *(A) Design of liquid-handling robotic reaction set up is shown. (B) The fitness of each unique physiochemical condition is represented as concentration of* n*-butanol after 24 h incubation at 30 ºC. Each bar corresponds to the physicochemical conditions in the order listed in **Table 6.1**. Error bars represent standard deviation of technical replicates with n=2. The black bar represents base case condition. Based on (B), the concentration of CoA was varied between zero and five mM and the concentration of NAD was varied between zero and ten mM. The concentrations after 24 h incubation at 30 ºC of (C)* n*-butanol, (D) lactate, and (E) acetate were measured. The average concentration of three technical replicates was graphed.*

We next set out to understand the reason for increased *n*-butanol yields. We found that the top 3 conditions corresponded to high levels of initial NAD, holding all other components constant, resulting in a statistically significant increase in *n*-butanol produced (p = 0.025). On the lower end, the four conditions that produced 0 mM *n*-butanol were a result of conditions with (1) no glucose added, (2) no NAD, and (3 & 4) high levels of CoA. This analysis not only served as an important control, because as expected *n*-butanol was not synthesized in the absence of glucose, but it also pointed to NAD as the most important added reagent positively correlated with high-level *n*-butanol production. Our results are consistent with the requirement of four molecules of NADH to produce one molecule *n*-butanol in the pathway (**Figure 6.1A**). Varying amounts of CoA is also a lever to control *n*-butanol metabolism, though it appeared from our initial screen that CoA levels would need to be finely tuned.

Given the freedom of design in adjusting cell-free system components, we further explored the synergistic effects of varying NAD (0-10 mM) and CoA (0-5 mM) concentrations, while keeping all other components constant, on *n*-butanol synthesis. This strategy allows us to rapidly map cofactor landscapes to explore local and global optima. We observed that pathway performance is carefully tuned, with an optimal balance between amounts of NAD and CoA needed to achieve the highest *n*-butanol values (**Figure 6.2C**). Our results suggest that when CoA and NAD are low, the flux through the acetyl-CoA node and the redox driving force to *n*-butanol are limited. When there is too much NAD or CoA present, flux likely is diverted to other nodes. To test this, we directly measured lactate (**Figure 6.2D**) and acetate (**Figure 6.2E**) production in the CFME reactions at 24 h. Lactate produced from pyruvate and acetate produced from acetyl-CoA, are two major by-products of cell-free *n*-butanol metabolism. We observed that acetate is inversely dependent on the starting concentration of CoA in the reaction and not dependent as much on NAD concentration. Acetate production is probably linked to CoA because the phosphotransacetylase and acetate kinase reactions convert acetyl-CoA to acetate and ATP while recycling phosphate. However, lactate has a similar optimal NAD to CoA ratio as *n*-butanol. Taken together, our results show that the cell-free framework and its barrier-free access to reaction conditions, is well-suited for rapidly acquiring physiochemical landscapes to assess and optimize biosynthetic pathway performance. This joins an emerging body of literature highlighting the value of cell-free systems for prototyping biological systems.[65,172,246-249]

6.4.2 Cell-free metabolic profiles are negatively affected by CFPS integration

Cell-free protein synthesis as a means of direct *in vitro* production of biosynthetic enzymes for crude-lysate metabolic engineering efforts (**Figure 6.1C**) enables faster DBT cycle times[173]. Unfortunately, we observed in our previous work that the yields of *n*-butanol in CFPS-ME reactions were lower than in CFME reactions (where biosynthesis enzymes are made in cells before extracts are made)[173]. We next set out to exploit the flexibility afforded by *in vitro* systems to modulate physiochemical conditions for systematically studying these differences and to demonstrate the ease of studying pathway performance in cell-free systems. Detailed quantification revealed how changes in glucose, lactate, acetate, and *n*-

butanol varied for CFME (**Figure 6.3A**) and CFPS-ME (**Figure 6.3B**) over the course of a 24 h reaction at 30 ºC. We observed three distinct regimes: (i) 0-3 h, (ii) 3-9 h, and (iii) 9+ h. During the first regime, there is a "lag" phase where glucose is consumed and metabolism starts up, making NADH. During the second regime, all metabolites are produced at a constant rate. The metabolite profiles level off during the third regime.
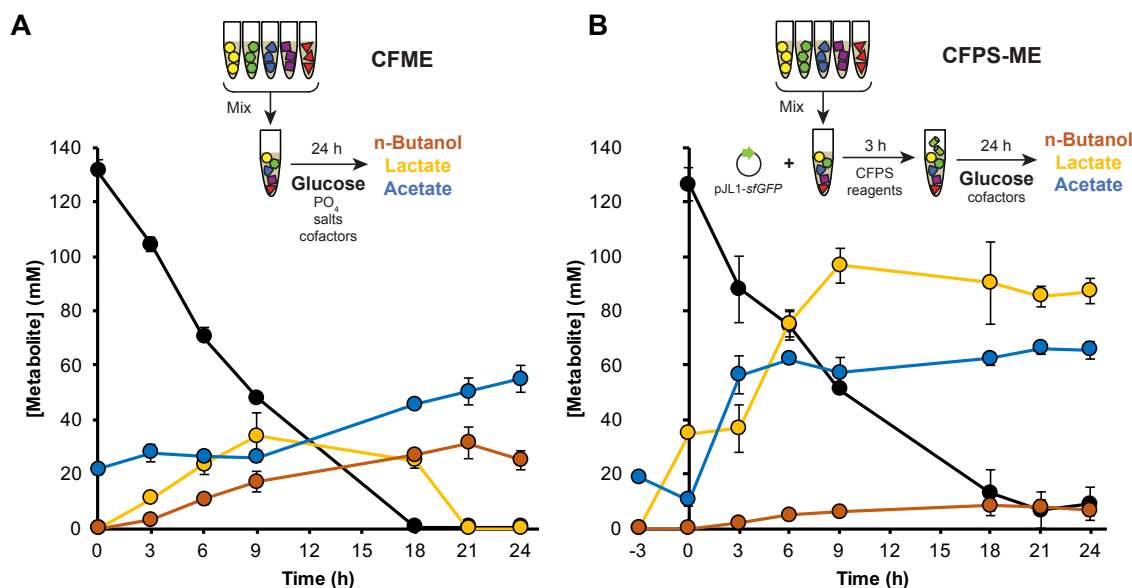
Figure 6.3. Protein synthesis alters cell-free metabolism and decreases n-butanol production. *Cell-free reactions were run for 24 h at 30 ºC and glucose (black), lactate (yellow), acetate (blue), and* n-*butanol (orange) were measured. (**A**) CFME reactions are run with five extracts mixed to have all* n-*butanol pathway enzymes present with glucose, NAD (3 mM), and CoA (1.5 mM). (**B**) CFPS-ME is run the same way with a 3 h incubation period with DNA and CFPS components prior to addition of glucose and CoA (1.5 mM). Each error bar represents standard deviations of technical replicates with n ≥ 3.*

Several differences between the cell-free metabolite profiles of CFME and CFPS-ME are notable. The two systems consume glucose at similar rates, but CFPS-ME systems produce *n*-butanol at ~25% of the titer of CFME systems. Additionally, lactate is produced and then consumed while acetate slowly increases after maximal lactate titer is achieved over the course of the CFME reaction. Whereas in CFPS-ME reactions, acetate and lactate are both only produced. Given the reduction in *n*-butanol yields, we wanted to understand the potential causes for the different metabolic profiles that arise when CFPS is used to make biosynthetic enzymes by directly accessing the physiochemical levers of the cell-free system. We know from our optimizations using liquid-handling robotics (**Figure 6.2**) that NAD and CoA are critical to

pathway performance, so it was not surprising that when we compare the original CFPS-ME system[173] (**Figure 6.3B**) with a CFPS-ME system containing the CFME-optimized NAD and CoA concentrations, we see an improvement in overall *n*-butanol production (**Figure 6.4**). However, increasing NAD concentrations still results in *n*-butanol production at ~80% of the titer of CFME systems. Therefore, we decided to further explore how CFPS might be affecting biosynthetic performance by directly accessing the physiochemical levers of the cell-free system.
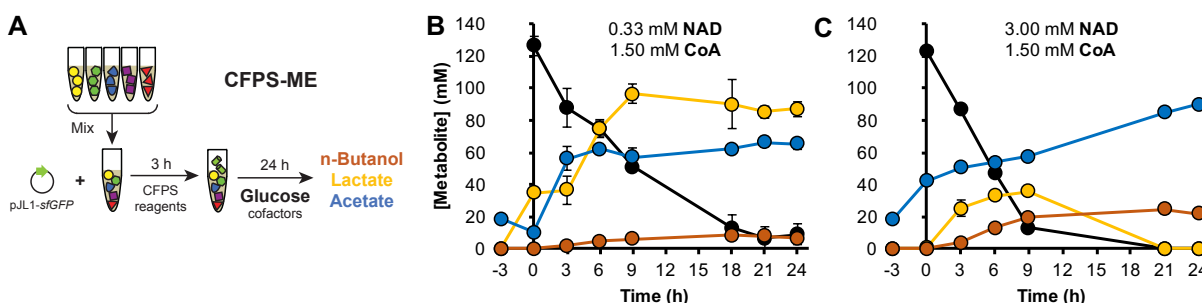
Figure 6.4. Increasing NAD concentrations for CFPS-ME results in n-butanol production at ~80% of the titer of CFME systems. *Cell-free CFPS-ME reactions were run with five extracts mixed to have all butanol pathway enzymes present. Each reaction was run with a 3 h incubation period at 30 ºC with DNA for sfGFP production and CFPS components prior to the addition of biosynthetic pathway starting substrates and cofactors and incubation for 24 h at 30 ºC. Time t = 0 h corresponds with the addition of glucose and cofactors. (**A**) The design of experiments is represented in schematic form. Glucose (black), lactate (yellow), acetate (blue), and butanol (orange) were measured at -3, 0, 3, 6, 9, 21, and 24 h. (**B**) CFPS-ME reactions were run with 0.33 mM NAD and 1.5 mM CoA. This figure is reproduced from main-text Figure 3B. (**C**) CFPS-ME reactions were run with 3 mM NAD and 1.5 mM CoA. Each error bar represents standard deviations of technical replicates with n ≥ 3.*

To produce proteins of interest, CFPS systems harness an ensemble of catalytic components necessary for energy generation and protein synthesis from crude lysates of cells. These activated catalysts act as a chemical factory to synthesize and fold desired protein products upon incubation with essential substrates, which include amino acids, nucleotides, DNA or mRNA template encoding the target protein, energy substrates, cofactors, and salts. We hypothesized that the differences we observed in **Figure 6.3** arose from either (i) changes in the extract that result from protein production or (ii) physiochemical components (which differ than those needed for biosynthetic pathway operation alone). To test which part was responsible, we performed CFME reactions without CFPS, with CFPS small molecule reagents alone, and with a complete CFPS system for sfGFP production prior to the addition of glucose and cofactors (activating *n*-butanol production) and incubation for 24 h at 30 ºC (**Figure 6.5**). The idea was to understand

the impact on cell-free metabolism with and without synthesis of a protein that would not affect the *n*-butanol biosynthesis pathway. We found that *n*-butanol is produced at the same rate and titer in both cases, suggesting that the CFPS reagents alone (or their metabolic byproducts) are most likely the causative factor to the observable metabolic differences.
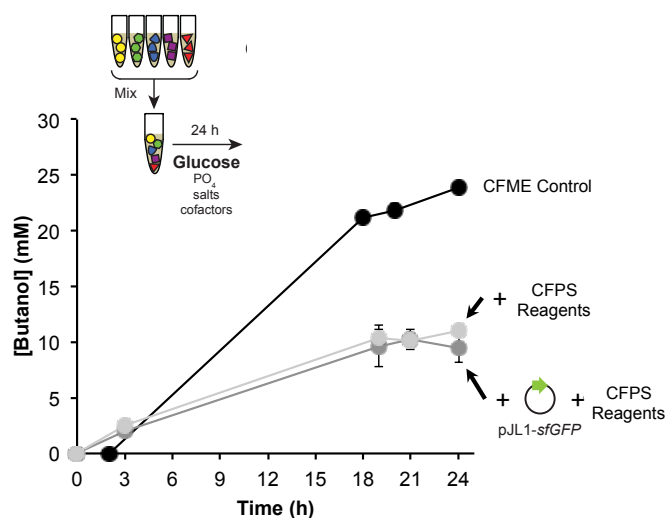
Figure 6.5. CFPS reagents rather than transcription-translation reactions are more likely the contributing factor to the metabolic differences between CFPS-ME and CFME. *Cell-free reactions were run with five extracts mixed to have all butanol pathway enzymes present with the addition of glucose and cofactors (at time t = 0 h) and incubation for 24 h at 30 ºC. Butanol is measured over 24 h. A control CFME reaction (black) is run as described here. Both CFPS-ME reactions were run with a 3 h incubation period at 30 ºC with (1) CFPS reagents alone (light grey) and with (2) DNA for sfGFP production and CFPS reagents (dark grey) prior to the addition of glucose and cofactors and incubation for 24 h at 30 ºC. Each error bar represents standard deviations of technical replicates with n ≥ 3.*

We then increased the resolution of our study by testing the effect of each of the small molecule CFPS components individually on metabolism (specifically the effect on glucose consumption as well as acetate, lactate, and *n*-butanol production) (**Figure 6.6A**). The small molecules unique to CFPS and not involved in CFME can be categorized into three classes: (1) components involved in transcription-translation (*e.g.,* amino acids, nucleotides), (2) components that mimic the cytoplasm (*e.g.,* putrescine, spermidine), and (3) components that modulate metabolism (*e.g.,* phosphoenolpyruvate (PEP), oxalate). We assessed the impact of supplementing CFME *n*-butanol synthesis reactions with components from each of these classes at increasing concentrations. The addition of transcription-translation substrates (nucleotides and amino acids, respectively) at 1X CFPS concentrations decreases *n*-butanol production by ~20% as compared to the standard CFME reaction (**Figure 6.6B**). The 'cytomimicry' components,

spermidine and putrescine, decrease *n*-butanol synthesis levels by ~18% at 1X CFPS concentrations and remain at these levels at higher concentrations of spermidine and putrescine (**Figure 6.6C**). In contrast, the metabolites involved in energy regeneration, oxalate and PEP, greatly reduce *n*-butanol production (**Figure 6.6D**). These trends are supported by corresponding levels of lactate and acetate (**Figure 6.7**). As a potent inhibitor of phophoenolpyruvate synthetase, we suspect that oxalate impacted the metabolic network around pyruvate and PEP to redirect flux resulting in lower levels of *n*-butanol. While each component tested influences *n*-butanol synthesis, during CFPS-ME these components exist at <0.5X concentrations during the reactions and likely do not individually cause drastic effects on *n*-butanol synthesis.
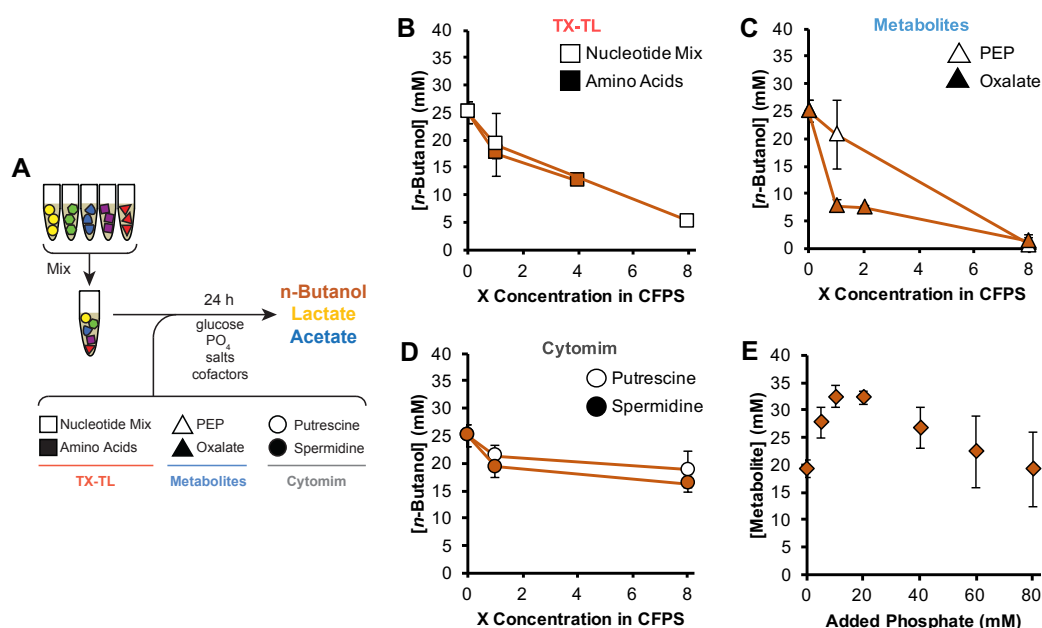
Figure 6.6. CFPS components rather than protein synthesis have a greater effect on n-butanol synthesis. *CFME reactions were run in the presence of individual CFPS reagents for 24 h at 30 °C. (A) The design of experiments is represented in schematic form.* n-Butanol was measured in the presence of varied (B) concentrations of transcription-translation specific reagents (TX-TL; nucleotide mix (white squares) and amino acids (orange squares)), (C) concentrations of CFPS energy metabolism reagents (Metabolites; PEP (white triangles) and oxalate (orange triangles)), (D) concentrations of CFPS reagents for cytoplasm mimicry (Cytomim; putrescine (white circles) and spermidine (orange circles)), and (E) K_2PO_4 concentrations (a proxy for phosphate generated during CFPS). Concentrations of CFPS reagents are varied as 'X Concentration in CFPS' where '1' would be the concentration of that reagent when a typical CFPS reaction is run. Each error bar represents standard deviations of technical replicates with n ≥ 3.*
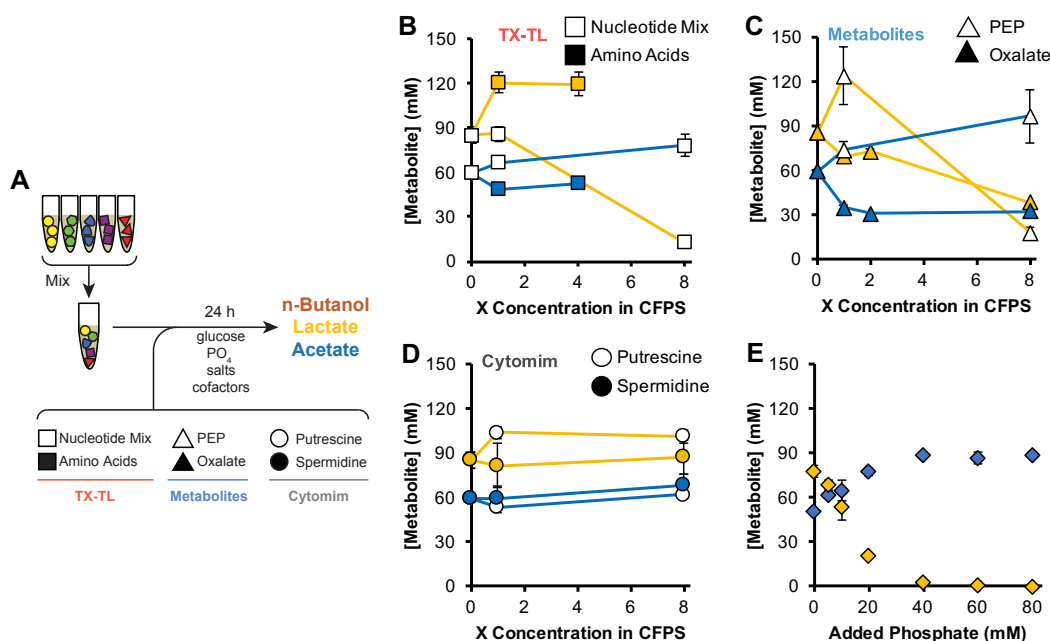
Figure 6.7. CFPS components rather than the protein synthesis process have a greater effect on lactate and acetate metabolism. CFME reactions were run in the presence of individual CFPS reagents for 24 h at 30 ºC. (A) The design of experiments is represented in schematic form. Lactate (yellow) and acetate (blue) were measured in the presence of varied (B) concentrations of transcription-translation specific reagents (TX-TL; nucleotide mix (white squares) and amino acids (colored squares)), (C) concentrations of CFPS energy metabolism reagents (Metabolites; PEP (white triangles) and oxalate (colored triangles)), (D) concentrations of CFPS reagents for cytoplasm mimicry (Cytomim; putrescine (white circles) and spermidine (colored circles)), and (E) K_2PO_4 concentrations (a proxy for phosphate generated during CFPS). Concentrations of CFPS reagents are varied as 'X Concentration in CFPS' where '1' would be the concentration of that reagent when a typical CFPS reaction is run. Each error bar represents standard deviations of technical replicates with n ≥ 3.

While PEP at <0.5X concentrations seems to not be detrimental, we know that PEP is used as a secondary energy source to regenerate ATP for energizing protein synthesis causing the concomitant release of free inorganic phosphate. We wanted to therefore test the effect of inorganic phosphate as we also know that phosphate can inhibit translation by sequestering magnesium[123] and could be deleterious to *n*-butanol synthesis as it does have an effect on glycolysis.[250] To test this hypothesis, we evaluated the addition of 0–80 mM inorganic phosphate in the form of potassium phosphate to our glucose-driven CFME *n*-butanol biosynthesis system at the start of the reaction (**Figure 6.6E**). We observed that *n*-butanol production has a phosphate optimum between 10 and 20 mM phosphate added to the system, but that high concentrations that would be representative of the CFPS-ME system (*i.e.,* > 30mM) were inhibitory relative to the optimum (~15% reduction). These observations re-enforce previous studies that have shown that

inorganic phosphate, when operating glycolysis, is a key component for optimizing cell-free systems.[51,250] Taken together, our analysis suggests each CFPS reagents together cause the reduction in *n*-butanol synthesis and that accumulated phosphate from PEP could be a major player in altering metabolism. By leveraging the ability to probe each physiochemical reagent afforded by the cell-free system we are able to parse individual and synergistic effects of physiochemical conditions on metabolism.

6.4.3 Non-phosphorylated energy sources decrease phosphate accumulation and shed light on the pH dependence of n-butanol synthesis

Readily altering physiochemical parameters gave us the ability to optimize *n*-butanol production in a cell-free system as well as observe differences between two cell-free systems (CFME and CFPS-ME) correlated with the addition of CFPS reagents. One result we observed was that phosphate is accumulated in CFPS-coupled systems (**Figure 6.6E**). This is due to the catabolism of PEP during CFPS. While the addition of PEP and each CFPS small molecule reagent contributes to productive protein synthesis, many alternative energy sources have been used to fuel CFPS such as glucose, pyruvate, and polysugars.[51,122,213,250-252] We hypothesized that alternative non-phosphorylated secondary energy sources for CFPS, instead of PEP, might mitigate the detrimental effect of phosphate on *n*-butanol metabolism. We tested five energy sources for CFPS during CFPS-ME each at varying metabolic nodes away from acetyl-CoA (PEP, pyruvate, glucose, maltose, and maltodextrin) (**Figure 6.8**). Unsurprisingly, PEP is the best energy source for protein synthesis, which makes sense because the physiochemical reagents in our CFPS system have been optimized for use with PEP. Additionally, PEP is the only source that accumulates >10 mM phosphate in the system as it is a direct phosphate donor. Further optimization of reagents could be useful for CFPS-ME using these alternative energy sources.
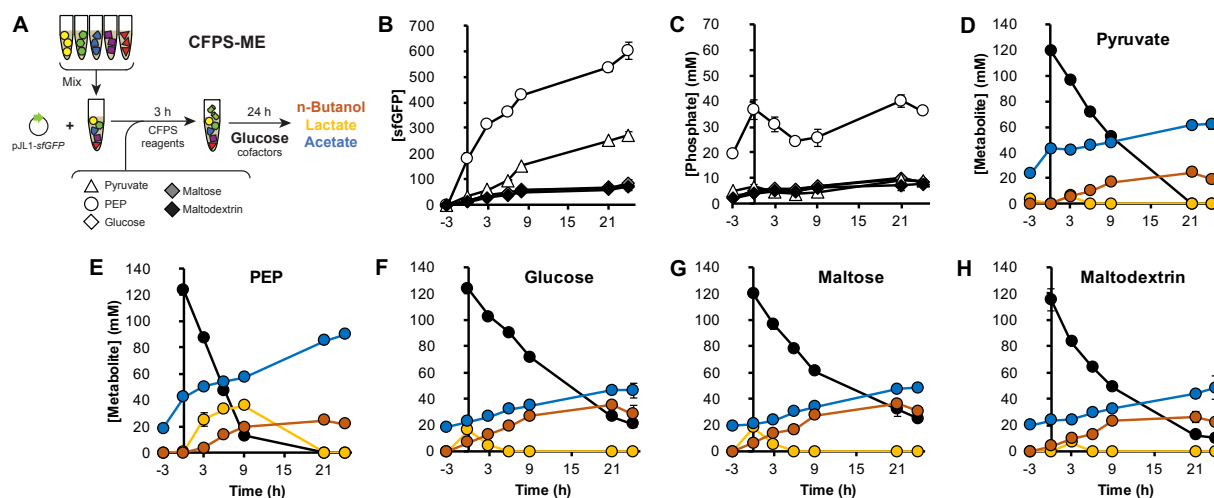
Figure 6.8. Using alternative energy substrates for CFPS decreases accumulated phosphate in CFPS-ME. *Cell-free CFPS-ME reactions were run with five extracts mixed to have all butanol pathway enzymes present. Each reaction was run with a 3 h incubation period at 30 ºC with DNA for sfGFP production and CFPS components prior to the addition of biosynthetic pathway starting substrates and cofactors and incubation for 24 h at 30 ºC. Time t = 0 h corresponds with the addition of glucose and cofactors. (A) The design of experiments is represented in schematic form. Glucose (black), lactate (yellow), acetate (blue), butanol (orange), phosphate levels, and sfGFP were measured at -3, 0, 3, 6, 9, 21, and 24 h. (B) sfGFP and (C) accumulated phosphate concentrations are measured for reactions testing pyruvate (white triangle), PEP (white circle), glucose (white diamond), maltose (grey diamond), and maltodextrin (black diamond) as biosynthetic pathway starting substrates. Metabolite (Glucose, lactate, acetate, butanol) profiles are shown for reactions testing (D) pyruvate, (E) PEP, (F) glucose, (G) maltose, and (H) maltodextrin as CFPS energy sources. Each error bar represents standard deviations of technical replicates with n ≥ 3.*

Being able to quickly adjust new reagents in cell-free systems, we next tested the same five energy sources as the sole substrate to fuel CFME *n*-butanol synthesis (**Figure 6.9A**). We observed that 33 mM pyruvate and PEP are unable to produce *n*-butanol (**Figure 6.9B-C**), likely because their conversion does not provide enough NADH driving force that would normally come from glycolysis to make *n*-butanol. Using 33 mM glucose, maltose, and maltodextrin produce *n*-butanol with similar profiles but at different rates (**Figure 6.9D-F**). This makes sense because each substrate is similarly catabolized during glycolysis, but maltose and maltodextrin must first be broken into glucose monomers. In addition, 33 mM maltose and maltodextrin are essentially twice as much carbon substrate as 33 mM glucose which likely plays a role in the difference between the *n*-butanol synthesis rates.
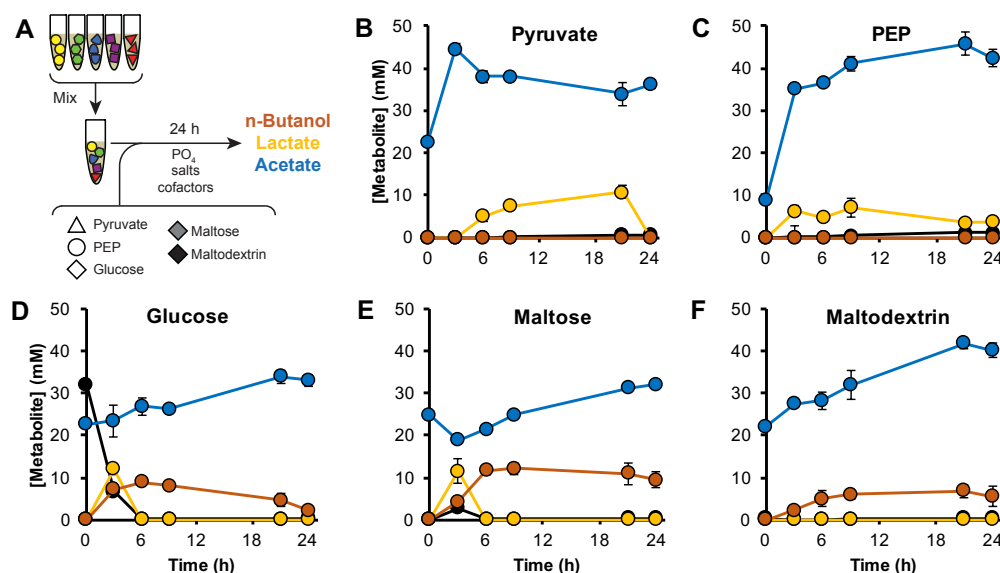
Figure 6.9. Cell-free CFME reactions can be run from alternative energy substrates to produce n-butanol. *(A) Cell-free CFME reactions were run with five extracts mixed to have all* n-*butanol pathway enzymes present with the addition of biosynthetic pathway starting substrates and cofactors and incubation for 24 h at 30 ºC. Glucose (black), lactate (yellow), acetate (blue), and* n-*butanol (orange) were measured at 0, 3, 6, 9, 21, and 24 h. Metabolite (glucose, lactate, acetate,* n-*butanol) profiles are shown for reactions testing 33 mM (B) pyruvate, (C) PEP, (D) glucose, (E) maltose, and (F) maltodextrin as biosynthetic pathway starting substrates. Each error bar represents standard deviations of technical replicates with n ≥ 3.*

We can also compare alternative energy substrates by observing accumulated phosphate and corresponding pH levels (**Figure 6.10**). Using PEP for CFME accumulates the most phosphate with pyruvate and maltodextrin also accumulating phosphate (**Figure 6.10B**). These systems all produce the least amount of *n*-butanol and further support the need to maintain low phosphate levels. Most interestingly, the pH of the system correlates with the amount of *n*-butanol that is produced (lower pH's correspond with higher *n*-butanol titers) (**Figure 6.10C**). This correlation could be due to low pH being beneficial for *n*-butanol synthesis or due to the increase in amount of energy metabolized (polysugar substrates) leading to increased *n*-butanol synthesis and thus lower pH. It is important to note that the compounds tested differ greatly in the amount of carbon supplied as substrate and metabolized as already mentioned. In addition, higher concentrations (120 mM) of glucose correlate with a decrease in pH indicated by black X's in **Figure 6.10C**, which makes sense as more glucose means four-times more carbon and therefore a more active metabolism. However, in congruence with previous reports of *Clostridium* metabolism, *n*-butanol production is optimal below pH of 7,[133,253,254] suggesting that the low pH may indeed be beneficial for *n*-butanol

synthesis. For CFME, using alternative buffers could alleviate limitations in *n*-butanol production. For CFPS, pH 7-8 is optimal for protein synthesis.[153] Together these results shed light on the essential physiochemical limitations of CFPS-ME for *n*-butanol synthesis only accessible by the physiochemical levers of crude-lysate systems. Moreover, they suggest that separating protein synthesis from *n*-butanol biosynthesis, a two-pot system, might improve pathway performance and its study.
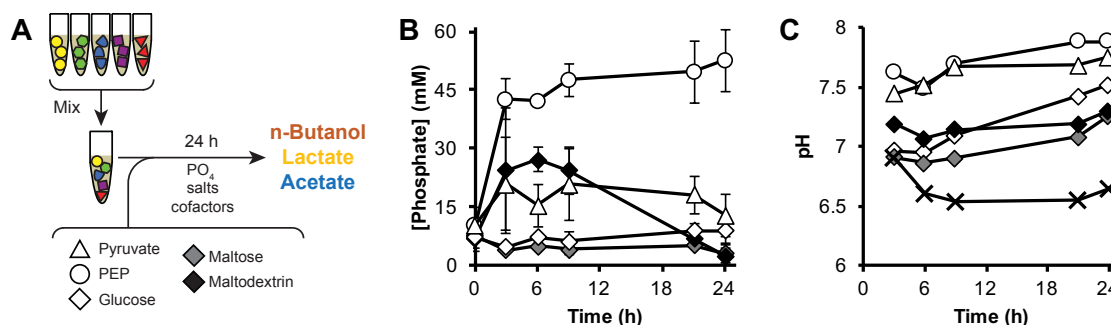
Figure 6.10. Using alternative energy substrates for CFME show varied operating phosphate and pH. *(A) Cell-free CFME reactions were run with five extracts mixed to have all* n-*butanol pathway enzymes present with the addition of biosynthetic pathway starting substrates and cofactors and incubation for 24 h at 30 ºC. Phosphate levels and pH were measured at 0, 3, 6, 9, 21, and 24 h. (B) Accumulated phosphate and (C) pH profiles are shown for reactions testing 33 mM pyruvate (white triangle), PEP (white circle), glucose (white diamond), maltose (grey diamond), and maltodextrin (black diamond) as biosynthetic pathway starting substrates. The pH of a CFME control reaction (120 mM glucose) is measured (black Xs) in (C). Each error bar represents standard deviations of technical replicates with n ≥ 3.*

6.5 Discussion

In this study, we present an approach to use crude extract-based cell-free systems to study biosynthetic pathway performance by physiochemical perturbations. First, we explored the backdrop of physiochemical parameters important for *n*-butanol production using liquid-handling robotics which enabled physiochemical landscapes to be mapped for optimizing pathway performance. This optimization improved cell-free *n*-butanol synthesis yields and demonstrates a generalizable strategy to inform biosystems design by modulating metabolism dynamics. Second, we compared and modified cell-free metabolic profiles in biosynthetic systems derived from (i) mix-and-match crude lysates where enzymes are made in cells prior to lysis (CFME) and from (ii) CFPS-coupled systems where enzymes are made post-lysis (CFPS-ME). This

taught us that the physiochemical complexity of an all-in-one CFPS-ME system inhibits *n*-butanol production and that physiochemical conditions play a major role in directing metabolism. Third, and to alleviate phosphate inhibition, we tested the impact of using non-phosphorylated energy substrates on cell-free pathway performance. We observed that CFME reactions fueled by PEP and pyruvate had higher accumulated phosphate levels and higher pH's (~8) as compared to reactions with glucose, maltose, and maltodextrin (~7). While this correlation between pH and *n*-butanol synthesis is noteworthy, the causal relationship is still unfounded. Whether using 33 mM maltose rather than 33 mM glucose, which is effectively twice as much carbon, would result in more *n*-butanol synthesis because of a lower pH or more *n*-butanol because of the increase in carbon supplied and catabolized is an open question. Luckily, our framework would allow manual adjustment of pH, carbon substrate, and by-products during reactions to test the causal nature of these types of relationships. Taken together, the modular nature of our approach is poised to facilitate multiplexed, automated study of biosynthetic pathways to inform systems design.

Having demonstrated the utility of modifying and controlling the physiochemical environment for cell-free biosynthetic systems, we interestingly found that lower system pH corresponded with increased *n*-butanol levels consistent with optimal pH's for *n*-butanol synthesis in native producers.[133,253,254] With this in mind and the flexibility to alter the physiochemical background, future work will focus on mimicking the cellular environment during different phases of growth to study pathway performance. For example, *Clostridium* undergo a biphasic metabolism during acetone-butanol-ethanol (ABE) fermentation. First the bacteria grow exponentially during the acidogenesis phase of metabolism. Then the bacteria enter stationary phase which corresponds with a switch to solventogenesis. These two phases represent two different metabolic profiles, pH's, and pathway operations. Just as we altered the physiochemical conditions of the cell-free systems to change the profiles of *n*-butanol, acetate, and lactate, one could imagine changing the physiochemical environment to mirror acidogenesis and solventogenesis to more accurately reflect pathway behavior. This could provide a more robust analysis of pathway dynamics to inform how pathways of enzymes operate during different bacterial growth regimes.

There are several applications of cell-free systems that the ability to control the physiochemical environment would facilitate. Most notably are the use of cell-free systems to study cellular metabolism, screen biosynthetic pathways for cellular metabolic engineering, and develop cell-free biomanufacturing

platforms. The tools and analysis performed in this work lays the groundwork to both rapidly prototype biosynthesis for *in vivo* use and to optimize pathways for *in vitro* use. Towards these applications it is important to have a robust and reliable cell-free setup. While there is variability of biosynthesis due to the accuracy of liquid-handling robotics, these systems are highly reproducible. Cell-free protein synthesis systems routinely perform well across lysates made from separate *E. coli* cultures.[169] Across the separate sets of mixed lysates (made from separate cultures of *E. coli*), the variability in *n*-butanol production after 24 h is insignificant (**Figure 6.11**). Cell-free systems would thus be able to hold up to the strains of high-throughput testing as well as stable biomanufacturing.
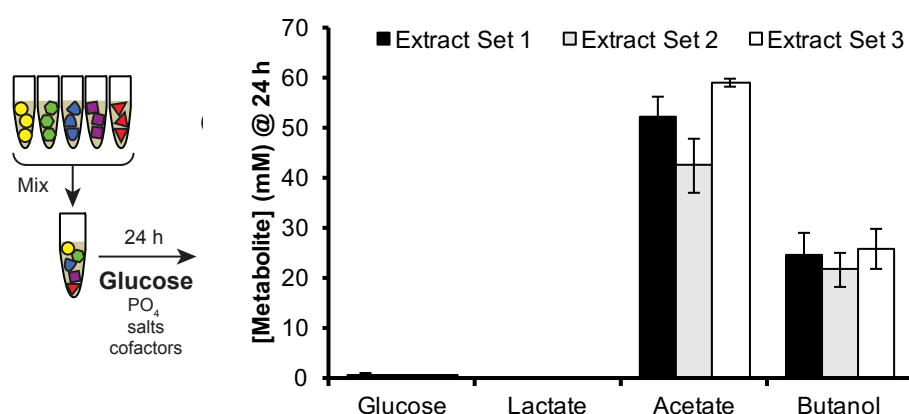
Figure 6.11. Metabolic production of n-butanol is reproducible across separate extracts. *Cell-free CFME reactions were run with five extracts mixed to have all butanol pathway enzymes present along with the addition of glucose, NAD, and CoA. Three separate biological replicates of* E. coli *were grown and lysed to make three sets of mixed extracts to compare the reproducibility of cell-free reactions. At 24 h, glucose, lactate, acetate, and* n-*butanol were measured. Each error bar represents standard deviations of technical replicates with n ≥ 3.*

Physiochemical manipulability makes cell-free metabolism highly controllable. We can use this feature to alter metabolism, tune and recover differences between cell-free biosynthetic systems, and potentially mimic different growth phase metabolisms and non-model organisms. While CFPS-coupled crude-lysate biosynthetic systems offer the best cell-free route to accelerate metabolic engineering DBT iterations, an important conclusion from our work is that there exists a dichotomy between the optimal physiochemical conditions for CFPS and those for biosynthetic pathway operation. This suggests that CFPS-coupled biosynthetic systems would be more powerful prototyping tools if multi-pot, phosphate-minimal, pH-controlled systems were developed. Looking forward, we expect to see the development of

cell-free biosynthetic systems developed as multiple "unit operations" that can be individually optimized and assayed gaining precise control over biosynthetic pathway construction. This concept of "unit operations", borrowed from chemical engineering, would offer even greater control over protein and metabolite synthesis allowing for better discernment of variation in metabolism due to changes in enzyme, making possible high-resolution strategies to prototype biosynthetic enzymes and pathways to speed up metabolic engineering DBT cycles.

6.6 Conclusions

This chapter provided a couple of key findings: (1) cell-free systems can rapidly probe metabolism by physiochemical perturbations and (2) optimal pH during CFPS contradicts optimal pH during *n*-butanol pathway operation. This work, as mentioned, also provided the insight to develop the system described in Chapter 7. This chapter is important because it is the first publication that goes into great detail as to how cell-free systems have an open reaction environment that we can control and regulate. This work also opens the door to controlling the physiochemical environment in ways that can mimic cellular metabolic environments.

Chapter 7: Building a bridge between cell-free experimentation and cellular metabolic engineering

If the only tool you have is a hammer, you tend to see every problem as a nail.
- Abraham Maslow

All previous chapters have led to the creation of this chapter. This chapter synthesizes all the developments of the *in vitro* prototyping and rapid optimization of biosynthetic enzymes approach. Going to conferences to talk about how cell-free systems can be used to build biosynthetic pathways, I always got ask "so what does that mean for me, the metabolic engineer who works inside living organisms?" It's a good question. A question that I was determined to answer or at least address in a meaningful way. Thus, we teamed up with LanzaTech, a leader in using microbes to capture carbon and re-use it for the production of biochemicals. This chapter describes the bridge between cell-free and cellular approaches—we can now screen biosynthetic pathways in cell-free systems to drastically reduce the number of pathway combinations that must be tried in cells. This becomes pertinent when dealing with non-model organisms like the microbes LanzaTech uses. At the risk of overselling this chapter, this work is really the climax of the dissertation, the fireworks of my PhD experience if you will. The chapter weaves together our state-of-the-art cell-free platform with the largest number of pathway combinations tried to date with machine learning to improve production of industrially relevant small molecules in live *Clostridium*.

7.1 Abstract

Industrial biotechnology remains an attractive, sustainable technology to compliment traditional efforts to produce chemicals to meet global demand. Industrially-relevant, non-model organisms such as

Clostridium provide excellent chassis for biochemical production from a diverse array of possible feedstocks but lack a robust toolset for manipulation to access of exotic molecules. Cell-free systems provide many advantages for accelerating design-build-test cycles due to the open reaction environment allows direct monitoring and manipulation of the system to study pathway performance. To date no attempts have been made using cell-free prototyping to improve engineering of industrially-relevant, non-model organisms. In this work, we report a new in vitro prototyping and rapid optimization of biosynthetic enzymes approach (termed iPROBE) to inform cellular metabolic engineering. The foundational principle is that we can construct discrete enzymatic pathways through modular assembly of cell lysates containing enzymes produced by cell-free protein synthesis rather than by living organisms. A key conceptual innovation is that the DBT unit can be cellular lysates rather than genetic constructs, allowing us to perform DBT iterations without the need to re-engineer organisms. We also develop a quantitative metric that combines, through multiplication, titer at reaction completion, rate during the most productive phase of pathway operation, and enzyme expression as measured by protein solubility (TREE score) to address the limitation of there being no easy method of analysis to provide informative bridging of cell-free data to cellular metabolic engineering. We demonstrate iPROBE and the use of the TREE score for the production of 3-hydroxybutyrate and n-butanol. We tested over 40 and 205 different enzyme combinations for 3-hydroxybutyrate and n-butanol production, respectively, more than any other metabolic engineering study and identify pathway combinations that produce at high-titers in vivo in Clostridium. We expect that iPROBE will accelerate design-build-test cycles for industrial biotechnology.

7.2 Introduction

For decades, scientists and engineers have turned to biological systems to help meet societal needs in energy, medicine, and materials—especially when chemical synthesis is untenable (e.g., antimalarial drugs). Often, biologically-produced small molecules are insufficient for production at the optimal titer, rate, or yield because natural sources are difficult to optimize and may not scale easily (e.g., plants grow slowly). Thus, engineers seek to design enzymatic reaction schemes in model microorganisms to meet manufacturing criteria.[4] Success in these endeavors depends upon identifying sets of enzymes

that can convert readily available molecules (e.g., glucose) to high-value products (e.g., medicines), with each enzyme performing one of a series of chemical modifications. Unfortunately, this is difficult because design-build-test (DBT) cycles—iterations of re-engineering organisms to test new sets of enzymes—are detrimentally slow due to the constraints of cell growth.[10] There is a cellular resource balancing act between the engineer's objective and the cell's viability. As a result, a typical project today might only explore dozens of variants of an enzymatic reaction pathway. This is often insufficient to identify a commercially relevant solution because selecting productive enzymes using existing single-enzyme kinetic data has limited applicability in multi-enzyme pathways and consequently requires more DBT iterations.[255] This challenge is exacerbated in industrially-relevant, non-model organisms such as *Clostridium* which lack a robust toolset for manipulation.

Non-model organisms provide excellent chassis for biochemical production of exotic molecules from a diverse array of possible feedstocks. The development of tools to engineer plants have grown over the last decade paving the way to harness the natural plant advantages such as photosynthesis, complex transmembrane enzymes, and more. *Clostridium* have been used industrially in acetone-butanol-ethanol (ABE) fermentations in the early-to-mid 20th century because of their unique solventogenic metabolism to produce large amounts of solvents (i.e., acetone and butanol) but were phased out of use due to the success of petroleum. On the other hand, acetogenic *Clostridium* can robustly ferment on a variety of gases. However, these strains tend to not have the natural machinery to produce such solvents and the tools to engineer *Clostridium* are lacking in their maturity. While developing tools for engineering *Clostridium* is ongoing, discovering methods to speed up metabolic engineering DBT cycles for these organisms would accelerate the re-industrialization of such powerful organisms.

Cell-free systems provide many advantages for accelerating DBT cycles.[138,144,244] For example, the open reaction environment allows direct monitoring and manipulation of the system to study pathway performance. As a result, many groups have used purified enzyme systems to study enzyme kinetics and inform cellular expression: testing enzymatic pathway performance *in vitro*, down-selecting promising pathway combinations, and implementing those in cells.[44,46,144] Crude lysates are becoming an increasingly popular alternative to purified systems to build biosynthetic pathways because they inherently provide the context of native-like metabolic networks.[171-173] For instance, the Panke group has shown that DHAP can

be made in crude lysates and real-time monitoring can optimize production.[172] In addition, our group has shown that 2,3-butanediol,[121] mevalonate,[171] *n*-butanol,[173] and more complex products[174] can be constructed in crude lysates with high productivities (>g/L/h). Recent work from the Murray[256] and Freemont[257] groups have attempted to prove the capabilities of crude lysates to prototype biosynthetic pathways to inform cellular expression but fail to test more than one set of enzymes at a minimal number of expression conditions, the main benefit of using cell-free systems. To date there have been no conclusive attempts at using cell-free systems to prototype useful numbers of enzyme variants and pathway combinations, that which you cannot do inside of a cell, for improving engineering of cellular metabolisms. In addition, no attempts have been made using cell-free prototyping to improve engineering of industrially-relevant, non-model organisms.

To address these challenges, we report a new *in vitro* prototyping and rapid optimization of biosynthetic enzymes approach (termed iPROBE) to inform cellular metabolic engineering. The foundational principle is that we can construct discrete enzymatic pathways through modular assembly of cell lysates containing enzymes produced by cell-free protein synthesis rather than by living organisms (**Figure 7.1**). This provides an unprecedented capability to test hundreds to thousands of pathways by avoiding inherent limitations of cell growth and thus diminishing the reliance on single-enzyme kinetic data. A key conceptual innovation is that the DBT unit can be cellular lysates rather than genetic constructs, allowing us to perform DBT iterations without the need to re-engineer organisms. We demonstrate that iPROBE can quickly study pathway enzyme ratios, tune individual enzymes in the context of a multi-step pathway, screen enzyme variants for high-performance enzymes, and discover enzyme functionalities. The rapid ability to build pathways in vitro using iPROBE allows us to generate large amounts of data describing pathway operation under several operating conditions. We also develop a quantitative metric that combines, through multiplication, titer at reaction completion, rate during the most productive phase of pathway operation, and enzyme expression as measured by protein solubility (TREE score) to address the limitation of there being no easy method of analysis to provide informative bridging of cell-free data to cellular metabolic engineering. By reducing the complexity of available cell-free data to one value we can now quickly screen and rank pathways in the cell-free environment and provide useful information for cellular metabolic engineering. We demonstrate iPROBE and the use of the TREE score for the production of 3-

hydroxybutyrate and *n*-butanol in *Clostridium*, an industrially relevant non-model organism. We tested over 40 and 205 different enzyme combinations for 3-hydroxybutyrate and *n*-butanol production, respectively, more than any other metabolic engineering study and identify pathway combinations that produce at high-titers *in vivo*.
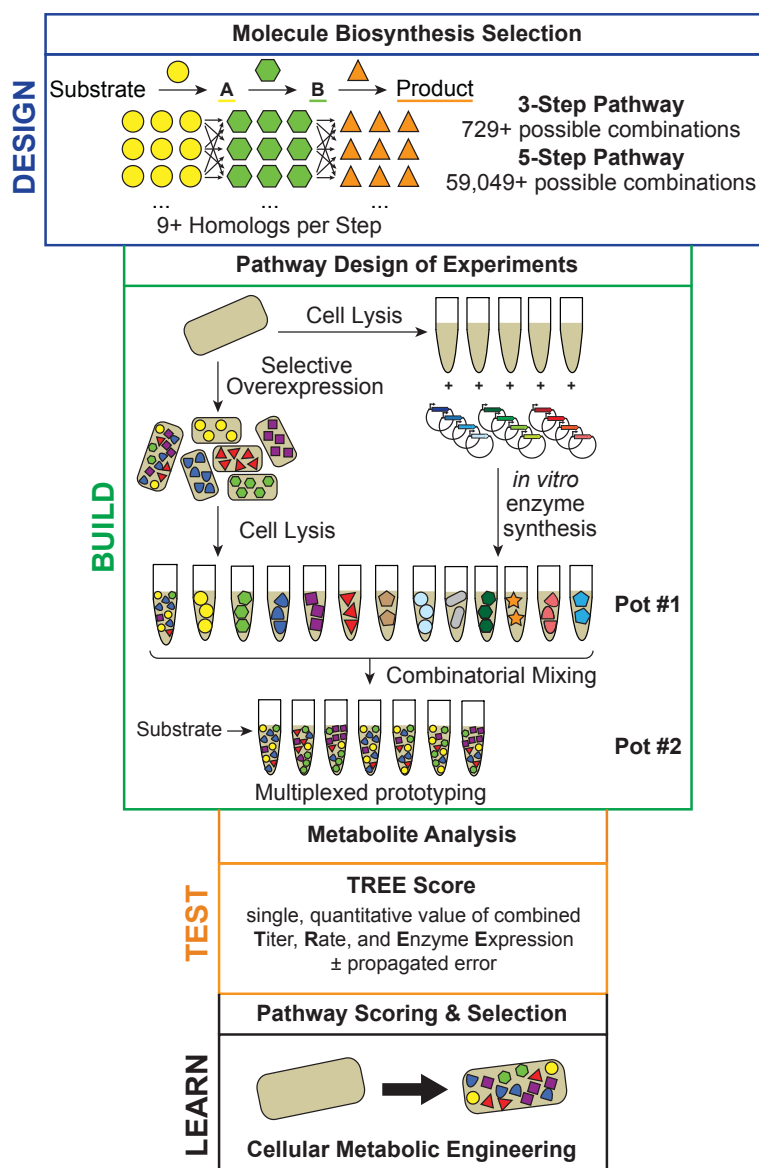
Figure 7.1 A two-pot cell-free framework for in vitro prototyping and rapid optimization of biosynthetic enzymes (iPROBE). *Schematic overview of the iPROBE approach following a design, build, test, learn framework. In the design phase, reaction schemes and enzyme homologs are selected. In the build phase, lysates are enriched with pathway enzymes via overexpression prior to lysis or by cell-free protein synthesis post lysis. Then, lysates are mixed to assemble enzymatic pathway combinations of interest. In the test phase, metabolites are quantified over time and data is reduced into a single quantitative metric for pathway combination scoring and selection. In the learn phase, cell-free pathway combinations are selected and implemented in cellular hosts.*

7.3 Materials and Methods

7.3.1   Bacterial strains and plasmids

Escherichia coli BL21(DE3) (NEB) was used for preparation of cell extracts which were used to express all exogenous proteins in vitro.[258] *Clostridium autoethanogenum* was used for *in vivo* characterization and fermentations (LanzaTech). DNA for all enzyme homologs tested were codon adapted for E. coli using IDT Codon Optimizer and for C. autoethanogenum using a LanzaTech in-house codon optimizer. E. coli adapted sequences are listed in **Table 0.5**. The cell-free plasmid, pJL1 (Addgene #69496), was used.

7.3.2   Cell Extract Preparation

E. coli BL21(DE3) cells were grown, harvested, lysed, and prepared using the detailed protocol in Chapter 4.

7.3.3   iPROBE Reactions

Cell-free protein synthesis (CFPS) reactions were performed to express each enzyme individually using a modified PANOx-SP system described in previous pubications.[152,153] Fifty-µL CFPS reactions were carried for each individual enzyme in 2-mL microcentrifuge tubes. Enzyme concentrations in CFPS reactions were quantified by 14C-leucine incorporation during in vitro translation. Then reactions performed for identical enzymes were pooled together when multiple reaction tube-volumes were needed to keep a consistent 50-µL reaction volume and geometry for every CFPS reaction. Based on molar quantities of exogenous enzymes in each CFPS reaction determined by radioactive measurement, CFPS reactions were mixed to assemble complete biosynthetic pathways in 1.5-mL microcentrifuge tubes. CFPS reactions constitute 15 µL of a 30-µL-total second reaction. When the total CFPS reaction mixture constituted less than 15 µL, 'blank' CFPS reaction was added to make the total amount of CFPS reaction up to 15 µL. The 'blank' reactions consist of a typical CFPS reaction with no DNA added. This 15 µL CFPS mixture was then added to kanamycin (50 µg ml−1), 120 mM glucose, magnesium glutamate (8 mM), ammonium glutamate (10 mM), potassium glutamate (134 mM), glucose (200 mM), Bis Tris (100 mM), NAD (3 mM), and CoA (3

mM; final reaction concentrations are listed. Reactions proceeded over 24 h. Measurements from samples were taken at 0, 3, 4, 5, 6, and 24 h.

7.3.4 Quantification of Protein Produced *in vitro*

CFPS reactions were performed with radioactive ^{14}C-Leucine (10 µM) supplemented in addition to all 20 standard amino acids. We used trichloroacetic acid (TCA) to precipitate radioactive protein samples. Radioactive counts from TCA-precipitated samples was measured by liquid scintillation to then quantify soluble and total yields of each protein produced as previously reported (MicroBeta2; PerkinElmer)[51,152].

7.3.5 Metabolite Quantification

High-performance liquid chromatography (HPLC) was used to analyze 3-HB and *n*-butanol. We used an Agilent 1260 series HPLC system (Agilent, Santa Clara, CA) via a refractive index (RI) detector. Analytes were separated using the Aminex HPX-87H and Fast Acids anion exchange columns (Bio-Rad Laboratories) with a 5 mM sulfuric acid mobile phase at 55 °C and a flow rate of 0.6 ml min^{-1}.

7.3.6 TREE Score Calculations

The TREE Score is calculated by multiplying the titer by the rate by enzyme expression metric.

$$TREE\ Score = Titer\ \cdot Rate\ \cdot (Average\ Solubility + [Total\ Enzyme]^{-1})$$

The titer is the metabolite concentration in the cell-free reaction at 24 h, when the reaction is complete. The error associated with titer is the standard deviation of reaction triplicates. The rate is the slope of the linear regression of metabolite concentrations taken at 3, 4, 5, and 6 h time points. The rate-associated error is the standard error of the slope calculated by the linear regression. The average solubility is calculated using measurements of soluble and total fractions of proteins, and their associated standard deviations (n=3), from 14C-leucine incorporation experiments for each enzyme. The error associated with solubility is the propagated error. The concentration of total enzyme is calculated by the addition of the final concentrations of each enzyme in the reaction with propagated error.

7.3.7 Design-of-experiments using neural networks

A neural network-based approach was used to design experiments to explore the large landscape of possible experiments because it is a data-driven method capable of identifying beneficial pathway

combinations without needing to understand the complex relationships between pathway combinations and their production capabilities. Parameters were optimized using multiple built-in functions for Python and MATLAB based on back propagation.

$$O_{nm} = F(x), \; where \; x = b_{nm} + \sum_{i=0}^{M} w_{(n-1)i} O_{(n-1)i}$$

$$and \; F(x) = \frac{1}{1 + e^{-x}}$$

7.4 Results

The new *in vitro* prototyping and rapid optimization of biosynthetic enzymes approach (termed iPROBE) to inform cellular metabolic engineering was built upon the foundation provided by our previous works. These works were the first time cell-free protein synthesis was used for biosynthetic pathway assembly with the benefit of an all-in-one-pot reaction. However, there are two major limitations: this reaction set-up relied on transcriptional and translational control to determine enzyme concentration; and pathway operation and enzyme synthesis can have conflicting physiochemical environments and thus decreased biosynthesis yields. By separating enzyme synthesis and enzyme utilization into two separate parts we are able to (1) modularly control enzyme concentrations because exact amounts can be added to the biosynthetic reaction and (2) match performance, production rate, of traditional cell-free approaches (**Figure 7.2**). In our approach, cell-free cocktails for synthesizing target small molecules are assembled in a mix-and-match fashion from crude cell lysates selectively enriched with pathway enzymes (**Figure 7.1**). This approach reconstructs pathways in two steps where the first step is enzyme synthesis via cell-free protein synthesis and the second step is enzyme utilization via substrate and cofactor addition to activate small molecule synthesis. Because we can quantify the amount of enzyme we add to the second pot during biosynthesis, we can for the first time individually titrate enzymes in the context of an entire biosynthetic pathway to study enzymatic activity. This differs from what has been done previously with two-pot systems[174] because we stop protein synthesis with kanamycin to control the amount of enzyme present

and we control the amount of CFPS reaction added to the second pot and thus reduce the negative physiochemical effects[170] on biosynthesis.
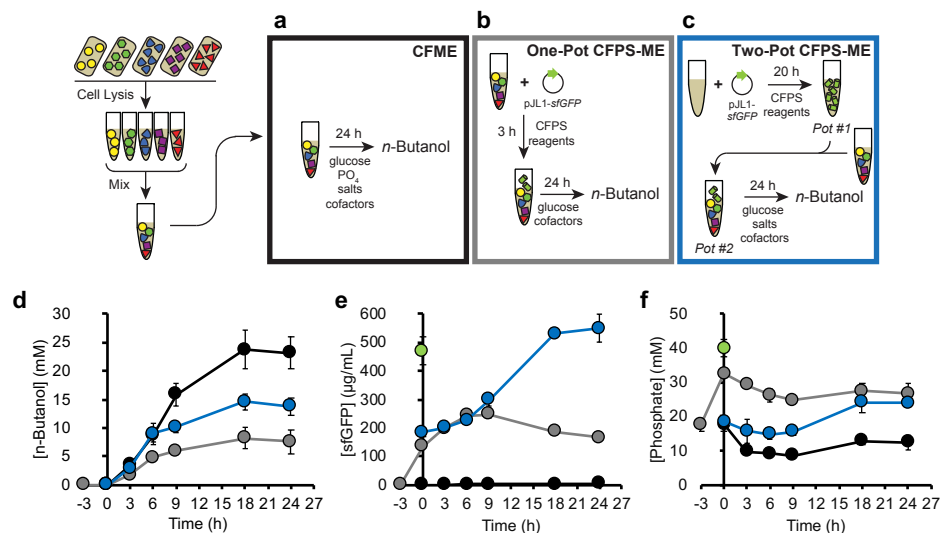
Figure 7.2. Two-pot cell-free pathway construction produces same linear rate as traditional CFME. *Schematic representations of (a) CFME, (b) one-pot CPFS-ME, and (c) two-pot CFPS-ME (iPROBE) are represented. (d) n-butanol production running CFME (black), one-pot CFPS-ME (gray), and two-pot CPFS-ME (blue) shows that two-pot CFPS-ME matches the production rate of CFME systems, while neither type of CFPS-ME systems matches final titers of CFME systems. (e) protein production measured by sfGFP made in each system is shown with a green circle representing CFPS alone. Two-pot CFPS-ME dilutes a CFPS reaction in and is shown in blue. (f) Inorganic phosphate levels are shown for each cell-free method. Two-pot CFPS-ME phosphate levels most resembles are two-pot system.*

In order to demonstrate the power and speed of the cell-free-centered design-build-test-learn cycle (**Figure 7.1**), we chose to use iPROBE to prototype 3-hydroxybutyrate (3-HB) biosynthesis, an industrially-relevant small molecule. We use a two, non-native enzyme pathway for the production of 3-HB (**Figure 7.3A**). This pathway requires the expression of a thiolase (Thl) and hydroxybutyryl-CoA dehydrogenase (Hbd). Our demonstration is composed in three parts: (1) we show that we can use cell-free experiments to generate design parameters for DNA construction of biosynthetic pathways in *Clostridium*; (2) we develop a single quantitative metric to score and select cell-free pathway combinations to inform cellular pathway selection; and (3) we screen several possible pathway combinations using cell-free experiments, select a subset of candidate cellular pathway combinations, and show cellular *Clostridium* 3-HB biosynthesis correlates with cell-free experimental results. iPROBE provides a quick and potent DBT framework to improve cellular metabolic engineering efforts.

7.4.1   iPROBE informs selection of genetic regulatory architectures in *Clostridium*

Expression of biosynthetic pathways in microorganism involves designing DNA constructs often involving a coding sequence, 5' and 3' regulatory elements, and vector maintenance components, among other parts. Selection of the optimal regulatory strengths (i.e., high, medium, low) for the expression of the coding sequence is an essential component to building a biosynthetic pathway in a cell. Several works have shown a tour-de-force in our ability to multiplex and increase the throughput in building libraries of regulatory architectures. However, as we find beneficial non-model organisms like *Clostridium* that we would like to engineer, transformation limitations prevent the extrapolation of such high-throughput techniques. Thus, we decided to use iPROBE to test specific enzyme concentrations to correlate with specific strength regulatory architectures in *Clostridium* allowing us to use cell-free systems to inform DNA construction.

In building a plasmid for *Clostridium* expression of Thl and Hbd for the production of 3-HB, we might construct them in one operon under one promoter and origin of replication. Changing the strength of the promoter and copy number of the plasmid changes the expression levels of Thl and Hbd. To mimic the expression of a single operon that we might build in *Clostridium* we built cell-free pathway combinations by co-titrating seven different enzyme concentrations of Thl and Hbd in our reactions. Specifically, we build seven cell-free reactions in increasing total concentrations added, combining CacThl and CklHbd at equimolar amounts (**Figure 7.3B-C**). We run each reaction for 24 h and measure the titer of 3-HB produced (**Figure 7.3D**). We see that as we increase the amount of enzyme add we increase the amount of 3-HB made up to a threshold amount as we approach 1 µM of each enzyme added. We then constructed plasmids expressing CacThl and CklHbd under 8 regulatory architectures (change in promoter strength and plasmid copy number) of increasing strength and transformed them into separate strains of Clostridium (**Figure 7.3E**). We ran small-scale fermentations of each strain under anaerobic conditions on carbon dioxide gas and measured stationary phase titers of 3-HB (**Figure 7.3F**). We can correlate cellular 3-HB production and corresponding plasmid regulatory strength with our cell-free 3-HB production and corresponding enzyme concentrations to build a cell-free to cell correlation table (**Table 7.1**). This now allows us to screen many different pathway combinations in cell-free systems and provide a rational recommendation for plasmid construction of those pathway combinations in *Clostridium*.
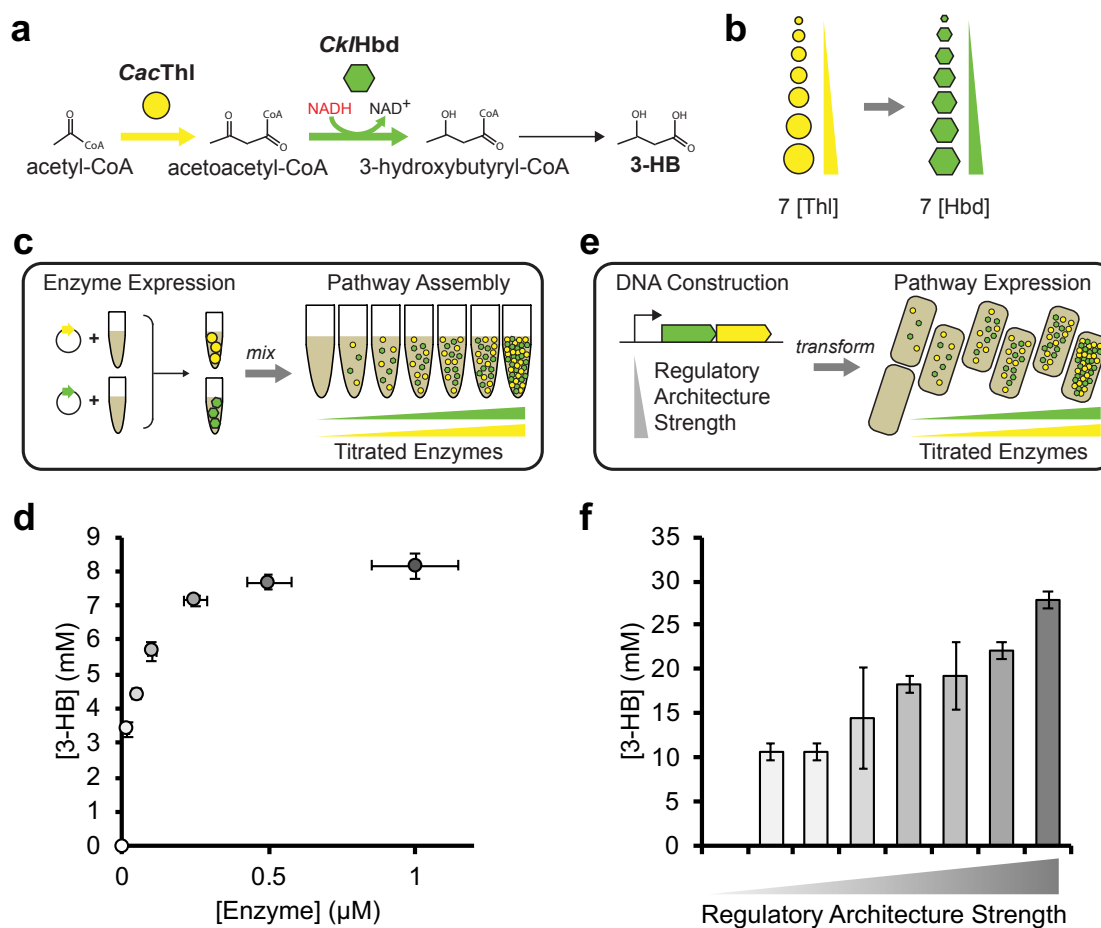
Figure 7.3. Enzyme concentrations can be tuned with iPROBE to inform genetic design for Clostridium expression of 3-hydroxybutyrate. *A reaction scheme for the production of 3-hydroxybutyrate is presented in panel (a). The design in (b) includes the co-titration of CacThl and CklHbd at seven concentrations (0, 0.02, 0.05, 0.1, 0.25, 0.5, and 1 µM). We built these seven designs in cell-free systems (c) by CFPS of each enzyme in separate lysates (Pot #1) followed by mixing to assemble full pathways for 3-HB production. (d) We measured 3-HB over the course of 24 h for each. We compared these results to Clostridium-based expression by building eight genetic constructs with varying promoters and plasmid copy number (e). (f) We measured final titer of 3-HB for each. Error bars represent technical triplicates. Error bars on enzyme concentrations are technical replicates with n > 3.*

Table 7.1. Expression recommendations for in vivo protein expression from cell-free enzyme assays.

Expression Levels	
in vivo	cell-free
low	< 0.1 µM
medium	0.1 - 0.3 µM
high	> 0.3 µM

7.4.2 Developing a metric to quantify biosynthetic pathway performance

Cell-free systems provide the luxury of being able to build pathway combinations (defined set of enzymes, or enzyme variants, and their concentrations) in hours to days. With iPROBE, we can now titrate in enzymes and observe activity from 10^2 to 10^3 enzyme combinations which is beneficial in defining the landscape of biosynthetic pathway performance. Characterizing pathway performance of this many combinations would be difficult in cells, especially *Clostridium* with long DBT turn-around times. However, rapidly analyzing large amounts of data generated from cell-free experiments can be difficult to decipher especially because which information of a metabolite time-course that is important to generate useful guidance for a scientist working in cells is unknown. In previous works, we quantified pathway performance by reporting final titer, but this does not take into consideration the value of a pathway combination that may perform much more quickly (a higher rate of production) than other combinations. We could also compare time courses to observe better-performing pathway combinations. However, visually it becomes difficult to discern differences with greater than 20 overlaid time courses. With the ultimate goal in mind being to prototype pathway combinations for rational engineering in cells, we need a better way of defining the success of a pathway combination in a cell-free experiment.

Titer, rate, yield (TRY), and being able to express active enzymes are important parameters for engineering a cell to express a biosynthetic pathway and defining its success. Pathway combinations with low values for any of these parameters could warrant the dismissal of those combinations for further characterization. Thus, we decided to develop a single, quantitative metric for our cell-free experiments that would incorporate these same parameters, titer, rate, and enzyme expression, as they pertain to the cell-free environment. Our titer, rate, and enzyme expression (TREE) score combines, through multiplication, titer at reaction completion, rate during the most productive phase of pathway operation, and enzyme expression as measured by protein solubility. Using our two-pot iPROBE reaction set-up, we designed five unique pathway combinations titrating different concentrations of Thl while maintaining a constant concentration of Hbd across each of the designs (**Figure 7.4B-C**). We measure 3-HB at 0, 3, 4, 5, 6, and 24 h for each of the five pathway combinations (**Figure 7.4D**). We then multiply our titer at 24 h (**Figure 7.4E**), the linear rate between 3 and 6 h (**Figure 7.4F**), and the sum of the average solubility of our enzymes,

Thl and Hbd (**Figure 7.4G**, black bar), and the inverse of the total enzyme concentration (**Figure 7.4H**) to

obtain our TREE score (**Figure 7.4I**) for each of the five pathway combinations. While the results of the

TREE score are not drastically different than looking at titer or rate alone, which they should not be, they

exaggerate differences that we might see in each component of the score. Being that we don't know if titer

or rate of cell-free systems are informative for cellular expression, combining them allows us to incorporate

both. Incorporating the difficulty of enzyme expression allows us to decrease scores of pathways that are

likely not to express well in cells. Being that we do not know *a priori* what component is the most

representative of pathway performance in a cell, this score allows us to incorporate each parameter that

could be influential. By reducing the complexity of available cell-free data to one value we can now quickly

screen and rank pathways in cell-free systems and provide useful information for cellular engineering.
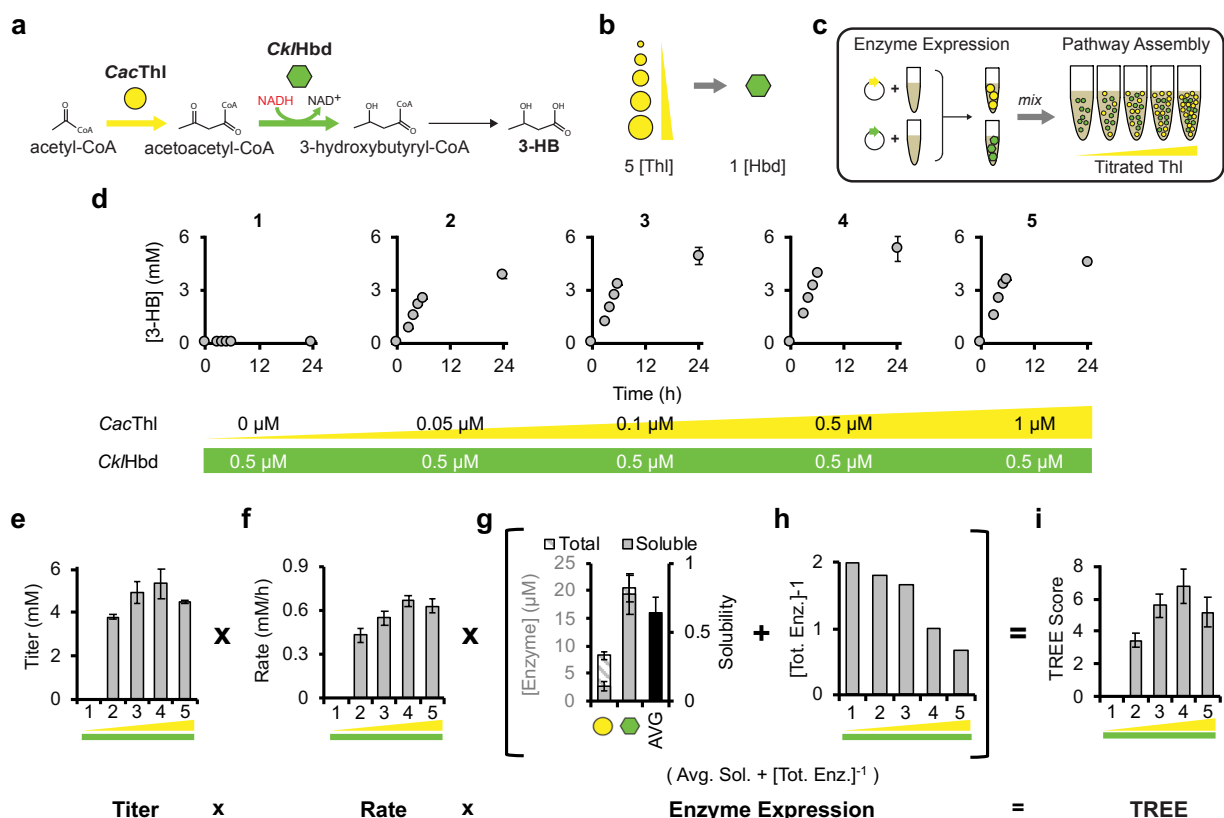
Figure 7.4. Individual pathway enzymes can be tuned in pathway context and scored using TREE scores with iPROBE. *(A) The pathway to produce 3-hydroxybutyrate from native metabolism (acetyl-CoA) is selected requiring two enzymes not natively present, CacThl and CklHbd. Five pathway combinations are designed to be built and tested varying the concentration of CacThl low to high while maintaining CklHbd at one concentration. (B) The five pathway designs are built by enriching two E. coli lysates with CacThl and CklHbd, respectively, by CFPS (Pot #1). Then, the five pathway combinations are assembled by mixing CFPS reactions containing CacThl, CklHbd, and no enzyme (blank) with fresh E. coli lysate. Kanamycin, to stop further protein synthesis, glucose and cofactors are added to start biosynthesis of 3-hydroxybutyrate. (C) 3-hydroxybutyrate is measured at 0, 3, 4, 5, 6, and 24 h after the addition of glucose for each of the five pathway combinations. Error bars are shown at 24 h and represent technical triplicates. From these measurements, 3-HB titer at 24 h and rate of production through 6 h is quantified. Enzyme expression is quantified by averaging the solubility of each enzyme and by the inverse of the total concentration of exogenous enzyme present. The TREE is score is then calculated for each pathway combination.*

7.4.3   iPROBE can inform the selection of pathway enzymes in *Clostridium*

With the TREE score in hand we decided to rapidly screen 36 pathway combinations, calculate

TREE scores, and measure expression in *Clostridium*, showcasing the iPROBE approach. To do this, we

test six enzyme homologs of each Thl and Hbd from different *Clostridium* species (**Figure 7.5A**). We

selected all simple pathway combinations of the 12 enzymes keeping a fixed total concentration of enzyme

added (high expression levels) to build in cell-free reactions (**Figure 7.5B-C**). By measuring 3-HB production over the course of 24 h, we are able to calculate TREE scores for each of the 36 pathway combinations (**Figure 7.5D**). We find that a majority of pathway combinations perform poorly with a handful achieving TREE scores above a value of 2. As engineering 36 separate strains of Clostridium is a non-trivial feat, we selected a subset of four pathway combinations to test in *Clostridium* labeled A, B, C, and D. We constructed and transformed DNA with strong regulatory architectures and each of the four-pathway enzyme sets into separate strains of *Clostridium* (**Figure 7.5E**). We ran small-scale fermentations of each strain under anaerobic conditions on carbon dioxide gas and measured 3-HB titers at four time points during the fermentation (**Figure 7.5F**). We observe that the best cell-free pathway combination (D) as determined by TREE score also performs the best of the four in *Clostridium* cells. The worst pathway combination in cell-free experiments (A) is also the worst performer in *Clostridium*. The other two pathway combinations (B, C) fall in the middle in cellular experiments. These data suggest that cell-free experiments are useful in selecting poor-performing pathways, pathway combinations that should not be tested in cells. In addition, scrutinizing differences in TREE scores of top performers may not be helpful in direct translation to cellular experiments. However, selecting top-performers in cell-free experiments to test in cells is useful. Through this demonstration iPROBE offers a framework to rapidly design, build, and test pathway combinations in cell-free experiments and learn how to implement pathway combinations with the best chances for metabolic engineering success in cells.

a

*Eco*Thl
*Cac*Thl
*Ckl*Thl
*Reu*PhaA/BktB
*Ckl*Thl2
*Cac*Thlm

*Cac*Hbd
*Cbe*Hbd
*Cac*Hbd2
*Cpa*Hbd
*Ckl*Hbd
*Csa*Hbd

NADH   NAD⁺

acetyl-CoA   acetoacetyl-CoA   3-hydroxybutyryl-CoA   **3-HB**

b

Screen Pathway
Combinations

c

Enzyme Expression   Pathway Assembly

mix

36 Pathway Combinations

d

TREE Score

iPROBE Pathways

A   B   C   D

e

DNA Construction   Pathway Expression

transform

f

[3-HB] (mM)

Time (au)

D   B   C   A
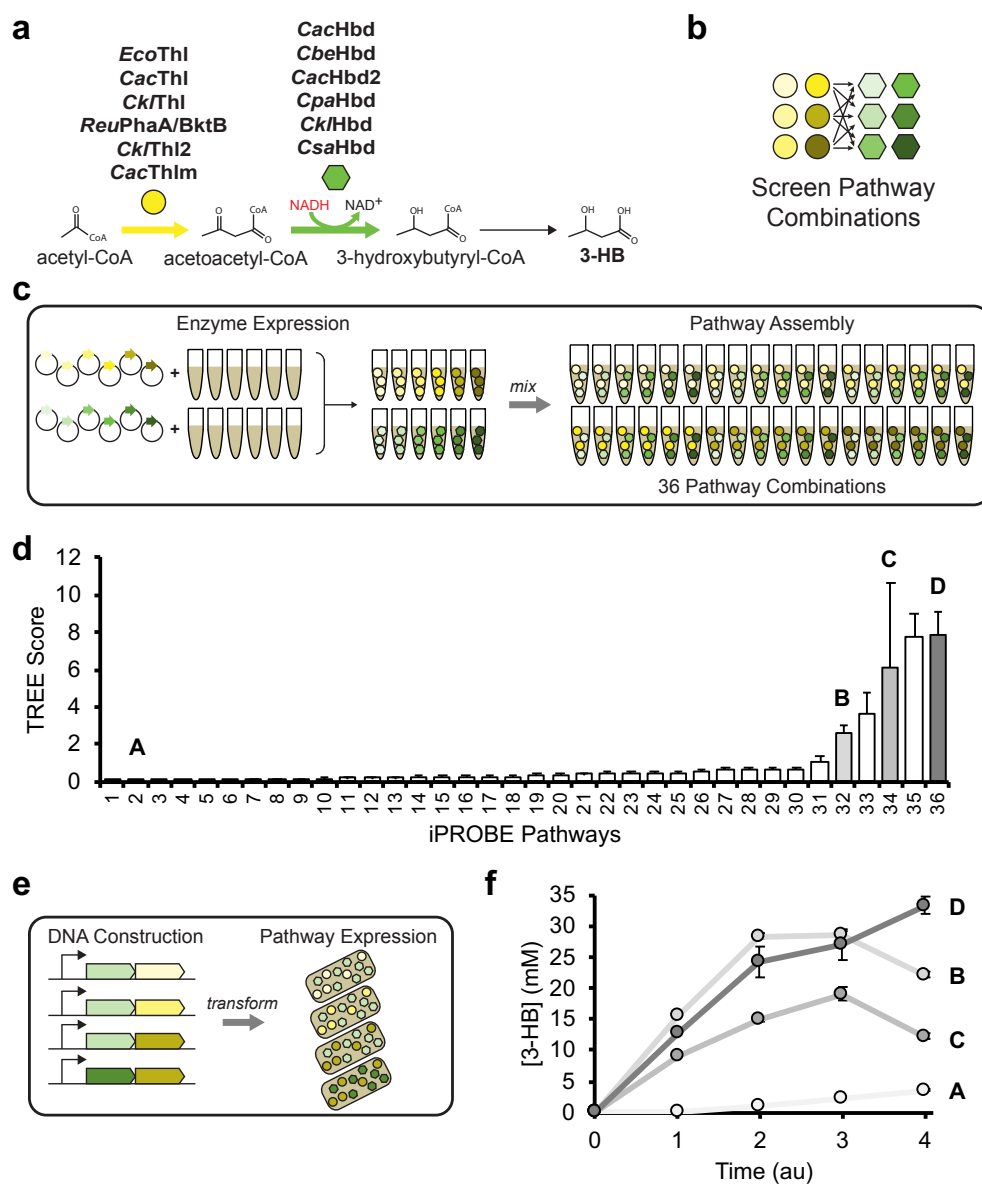
Figure 7.5. Enzymatic pathways can be screened with iPROBE to inform Clostridium expression for optimizing 3-hydroxybutyrate production. A reaction scheme for the production of 3-hydroxybutyrate is presented in panel (a). Six homologs have been selected for each reaction step. The design in (b) includes the testing of six Thl homologs and six Hbd homologs at 0.5 µM each. We built each possible combination in cell-free systems (c) constituting 36 unique pathway combinations. We built these cell-free pathways by expressing each of the 12 enzyme variants in lysates by CFPS. We then mixed each to try all 36 possible combinations keeping enzyme concentration fixed. (d) 3-HB was measured, and TREE scores were calculated and plotted for each iPROBE pathway combination. We then selected four pathway combinations to test in Clostridium (A, B, C, and D). These pathways were built in high copy plasmids with the highest strength promoters in single operons (e). Clostridium strains containing these pathway combinations were then fermented on gas and 3-HB was measured over the course of fermentation (f). Error bars represent technical triplicates.

7.4.4 Scaled-up fermentations of 3-hydroxybutyrate pathway prototype

Once pathway combinations are screened with iPROBE and subsets are tested in Clostridium, the optimal prototypes can be optimized for large-scale industrial fermentations. To show this, we chose our best selected 3-HB pathway combination in Clostridium. We scaled up fermentation of this strain in continuous fermentation (**Figure 7.6A**). Over the course of 16 days, we monitored 3-HB and biomass over a 2-week fermentation (**Figure 7.6B-C**). In our strain we observe high-titers of 3-HB, ~15 g/L. This is promising for the implementation of this strain in industrial fermentations. LanzaTech already has constructions in China that would allow the fermentation of Clostridium on waste gas streams to produce economic amounts of molecules such as 3-HB. By using iPROBE to down-select many possible pathway combinations we can develop promising industrial strains of Clostridium in a fraction of the time it took previously.
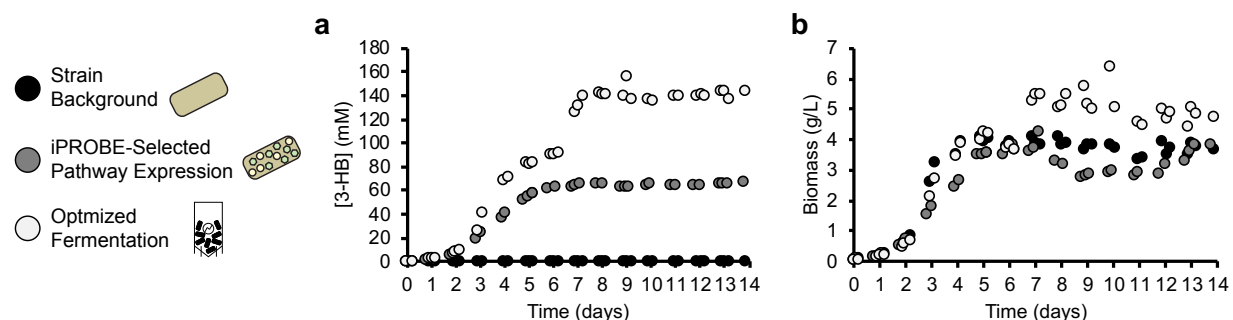
Figure 7.6. Clostridium fermentations show improved production of 3-hydroxybutyrate and n-butanol. The iPROBE-selected optimal pathway, CacThl and CklHbd1, for 3-HB production is built in a Clostridium strain and run in 14-day continuous fermentation with (a) 3-HB and (b) biomass monitored.

7.4.5 Cell-free pathway prototyping for *n*-butanol biosynthesis

The iPROBE framework can be used for larger pathways, such as the 6-step pathway from acetyl-CoA to *n*-butanol (**Figure 7.8A**). This pathway requires both a thiolase (Thl) and hydroxybutyryl-CoA dehydrogenase (Hbd) along with a crotonyl-CoA dehydrogenase (Crt), a trans-enoyl-CoA reductase (Ter), and a bifunctional aldehyde/alcohol dehydrogenase (AdhE). We selected one Thl, Hbd, Crt, and Ter to set as the 'base case' based on the enzymes used previously in our work (EcoThl, CbeHbd, CacCrt, and

TdeTer; **Figure 7.8A**, highlighted in blue). We first use this base case and try to modulate enzyme concentrations to improve butanol production. Testing 26 pathway combinations we quickly find that this approach does not achieve meaningful improvements (**Figure 7.7**). Just as we did with the prototyping of 3-HB, we decided to select new homologs to test for each step of the pathway.
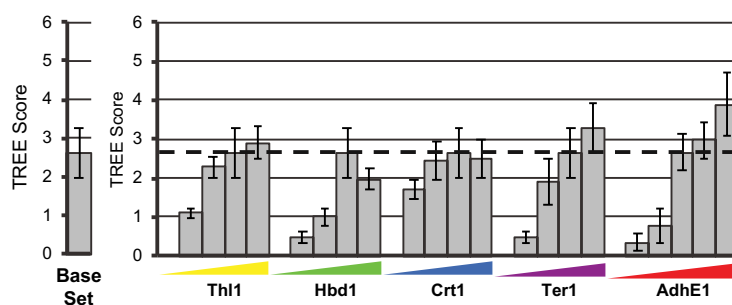
Figure 7.7. Titrating individual base case enzymes for the production of n-butanol. *Concentrations of 0.3 µM were used for EcoThl, CbeHbd, CacCrt, and TdeTer while 0.6 µM was used for CacAdhE for the Base Set and for the background enzyme combinations for each enzyme titration. TREE scores are calculated for each combination.*

We chose six homologs of each Thl, Hbd, Crt, and Ter to prototype with iPROBE for butanol production (**Figure 7.8A**). Because we can rapidly titrate in enzymes to study their pathway effectiveness without needing to re-engineer DNA constructs, we decided to test five concentrations of each enzyme homolog in pathway context consisting of our base case set of enzymes, totaling 120 pathway combinations, including those in **Figure 7.7** (**Figure 7.8B**). We built these combinations in cell-free reactions and calculated TREE scores for each (**Figure 7.8C**). The dashed line is placed at the TREE score resulting from the base case set of enzymes. A majority of the enzyme homologs, in the context of the base case, do not out-perform the original base case. This is not too surprising as the base case has been extensively characterized and tested throughout the literature.[134,136,158] However, we do find that CklHbd1 can double the TREE score at high concentrations and even out-performs the base case Hbd at lower concentrations. While testing as many concentrations and homologs as we did is beyond what would be tried in cells, exchanging one enzyme at a time in the base case set of pathway enzymes is typical. By doing this we potentially miss out on synergies that different sets of enzymes might provide us. Yet testing every possible enzyme set and concentration is simply not feasible.
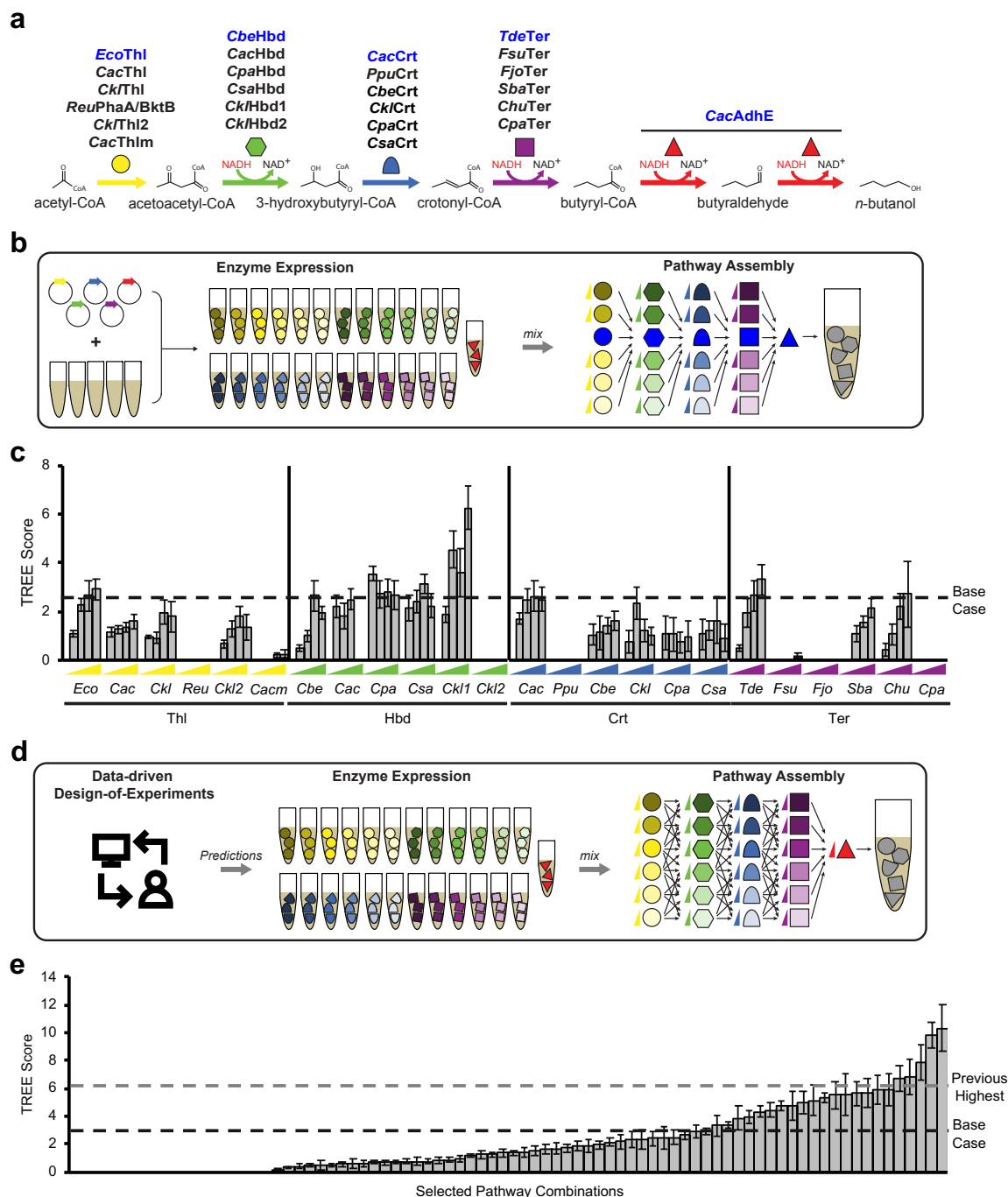
Figure 7.8. Cell-free pathway testing combined with machine learning quickly screens 205 unique pathway combinations for the production of n-butanol. *A reaction scheme for the production of n-butanol is presented in panel (a). Six homologs have been selected for each of the first for reaction steps and shown in panel (a). The strategy for running an initial set of reactions is to test each homolog individually with the base case set of enzymes (blue) at 5 concentrations. (b) These 120 pathway combinations are run in cell-free reactions according to the two-pot iPROBE methodology and (c) TREE scores are calculated from 24 h n-butanol time-courses. (d) Neural-network-based algorithms are used from the data presented in (d) to predict enzyme sets and concentrations. These are then built in cell-free systems. (e) TREE scores are calculated for all 205 pathway combinations tested in total from 24 h n-butanol time-courses.*

Harnessing the rapid capabilities of iPROBE we implemented a design-of-experiments using a neural-network-based algorithm to predict beneficial pathway combinations (enzyme sets and concentrations) in order manage the landscape of testable hypotheses (**Figure 7.8D**). Specifically, we trained the neural networks on pathway combinations in **Figure 7.8C** (i.e., inputs as enzymes and concentrations, outputs as titers and rates of the TREE score). Ten implementations of this algorithm were run to generate 100 pathway combination predictions to maximize TREE scores. Of these 100 predictions, combinations that suggested enzyme combinations <0.01 µM were ruled out. Of the remaining predictions 43 were selected and built in cell-free reactions (**Figure 7.9C**). In addition, we looked more closely at varying enzyme concentrations of the base set enzymes simultaneously (21 pathway combinations; **Figure 7.9A**) and hand-selected 18 pathway combinations based on our understanding of the biosynthesis (**Figure 7.9B**) to test with iPROBE. TREE scores were calculated for each of the 81 pathway combinations (**Figure 7.8E**). Nearly 19% of the 205 total pathway combinations screened in cell-free reactions have higher TREE scores than our base case (**Figure 7.8E**). Without having to test every possible combination, we were able to test a subset achieving ~4 times higher TREE scores over the base case pathway combination. Our neural network design was essential in achieving such high TREE scores, showcasing the importance of merging computational and experimental design. We could then proceed to use this information to inform cellular expression by providing the top designs based on our cell-free pathway combinations for expression in *Clostridium* (**Table 7.2**). Generating data for over 200 pathway combinations, built cumulatively in 12 days in cell-free reactions, would take quite a long time to generate in cells. This study shows the largest number of pathway combinations screened for a biosynthetic pathway to date and the benefits of such a screen.
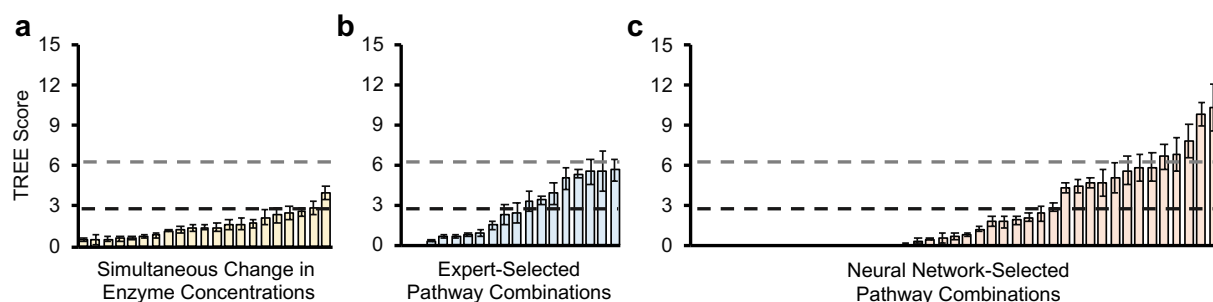
Figure 7.9. Cell-free experimental TREE scores for expert-selected and NN-based design of experiments. *TREE scores were calculated for pathway combinations experimentally tested with iPROBE for (a) simultaneous changes in each enzyme's concentration using the base case set of enzymes (21 combinations), (b) expert-selected pathway combinations based on data in Figure 5C and understanding of biosynthesis (18 combinations), and (c) pathway combinations selected from the neural network approach (43 combinations). TREE scores were calculated based on 24 h time-course data of n-butanol production.*

Table 7.2. Pathway designs selected for Clostridium expression based on iPROBE screening. *The first design is the base case. Designs 1 through 12 are the ranked highest performing pathways. Each enzymatic step has a homolog and expression level selected based on cell-free expression.*

	Thl		Hbd		Crt		Ter	
0	**Eco**	**high**	**Cbe**	**high**	**Cac**	**high**	**Tde**	**high**
1	Eco	high	Ckl1	high	Cac	low	Tde	high
2	Eco	high	Ckl1	high	Cac	high	Tde	medium
3	Eco	medium	Ckl1	high	Cac	low	Tde	high
4	Eco	high	Ckl1	high	Cac	high	Tde	high
5	Eco	medium	Ckl1	high	Cac	high	Tde	high
6	Ckl	medium	Ckl1	high	Cac	medium	Tde	high
7	Ckl	medium	Cac	high	Cac	medium	Tde	high
8	Cac	medium	Ckl1	high	Cac	medium	Tde	high
9	Cac	high	Cac	high	Cac	high	Tde	high
10	Eco	high	Ckl1	medium	Cac	high	Tde	high
11	Eco	medium	Ckl1	high	Cac	medium	Tde	high
12	Eco	high	Cbe	high	Cac	low	Tde	high

We stumbled into a roadblock as many of these pathway combinations resulted in DNA cloning failures due to toxicities of expressing *n*-butanol pathway enzymes. Thus, we chose to clone a subset that would provide insights into pathway expression in *Clostridium*. Naturally, solventogenic clostridia that produce *n*-butanol use an alternative to Ter, Bcd, which requires ferrodoxin-dependent electron transfer flavin (EtfAB) proteins. Ter enzymes are thus not used in *Clostridium*, but rather only in *E. coli*. We wanted to see if we could use Ter to produce *n*-butanol in *Clostridium*. We built plasmids to express CacThl, CklHbd1, CacCrt, and each of the Ter enzymes (TdeTer, FsuTer, FjoTer, and SbaTer) under four regulatory

strength (**Figure 7.10G**). We find that in our cell-free reactions, TdeTer and SbaTer produce butanol, while FsuTer and FjoTer are ineffective (**Figure 7.10H**). This is the first time a trans-enoyl-CoA reductase has been used in *Clostridium* to produce *n*-butanol. We proceeded to exchange our NADH-dependent pathway enzymes (Thl, Hbd, and Crt) with NADPH-dependent equivalents used previously in the literature (PhaA, PhaB, and PhaJ) because *Clostridium* species can prefer NADPH. By doing this we are able to increase *n*-butanol production ~4x (**Figure 7.10I**). By exchanging TdeTer with the CacBcd-EtfAB complex we are able to further increase *n*-butanol production by an order of magnitude. This suggests that as we move forward using *E. coli*-based iPROBE to screen pathways for expression in non-model organisms like *Clostridium*, we must take into consideration both the redox and other organism uses of the cells we'd like to engineer.
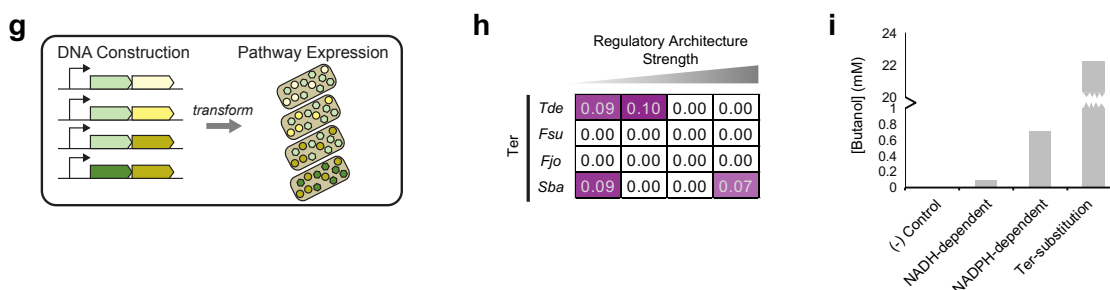
Figure 7.10. In vivo characterization of n-butanol pathway combinations. *(g) Pathways were built in plasmids with varied strength copy numbers and promoters in single operons. (h) Clostridium strains containing pathway combinations using CacTHL, CklHbd1, and CacCrt with TdeTer, FsuTer, FjoTer, and SbaTer were then fermented on gas and n-butanol was measured over the course of fermentation. (i) Pathways with (1) no enzymes; (2) CacTHL, CklHbd1, CacCrt, and TdeTer; (3) PhaA, PhaB, PhaJ, and TdeTer; and (4) PhaA, PhaB, PhaJ, and Bcd-EtfAB were run in fermentations and n-butanol was measured.*

7.5 Summary

We demonstrate a new framework, iPROBE, that incorporates a highly controllable and modular cell-free platform for constructing biosynthetic pathways with a novel quantitative metric for pathway performance selection to engineer and improve small molecule biosynthetic in non-model organisms that can be arduous to manipulate. Specifically, we show that by screening biosynthetic pathway combinations for the production of 3-hydroxybutyrate and *n*-butanol in cell-free reactions, combined with machine-

learning algorithms, we can rationally select a handful of pathways to construct and test in *Clostridium*. We test over 40 pathway combinations for the production of 3-HB and over 200 pathway combinations for the production of *n*-butanol, the largest set of pathway combinations in one study in the literature. Through this approach we engineer a strain of *Clostridium* that produces high-titers and yields of 3-HB large-scale continuous fermentations. We anticipate that iPROBE will facilitate rapid design-build-test cycles of biosynthetic pathways, decrease the number of the strains that need to be engineered, and in turn increase the number of industrially-viable biosyntheses.

7.6 Acknowledgments

This work would not be possible without a couple key collaborations. Michael Köpke at LanzaTech along with his team, Alex Juminaga, Rasmus Jensen, and Yongbo Yuan in particular, provided all Clostridium work as well as advice on what would be useful information from cell-free experiments. In addition, the machine learning aspects of this project arose from a collaboration with Lockheed Martin through AFRL with Jeff Stuart and Dave Coar. In addition, Quentin Dudley was integral in ideation of the prototyping paradigm, iPROBE.

7.7 Conclusions

This chapter focused on using cell-free systems to improve metabolic engineering for industrially relevant small molecule production in living cells. We develop a single quantitative metric, the TREE score, that allows us to rank screened pathway combinations in cell-free systems by parameters that we know are meaningful in living organisms. We then screen pathways for 3-hydroxybutyrate and *n*-butanol production in cell-free systems, one of the largest studies of biosynthetic pathways in the literature and that the screening leads to large-scale continuous fermentation of *Clostridium* to make these molecules. This chapter is the culmination of all previous chapters and highlights the capacity cell-free systems have for prototyping biosynthetic pathways.

Chapter 8: Natural product discovery

Mistakes are, after all, the foundations of truth, and if a man does not know what a thing is, it is at least an increase in knowledge if he knows what it is not.

- Carl Jung

This chapter stems from a personal interest in antibiotics and tough problems. While antibiotics do not appear anywhere in this chapter, they are natural products made from large multimodular enzymes—those do appear all throughout this chapter. This chapter was a side project throughout the last three years of my PhD and ends up being a series of vignettes of the difficulties and successes of expressing large non-ribosomal peptide synthetases and their products. This is challenging because expressing large proteins in cell-free is not common and there is not a lot of literature describing what to do troubleshoot such reactions. This chapter attempts to catalog my trials of expressing several large enzymes.

8.1 Abstract

Medicines are often derived from fungal and bacterial natural products. Many of these molecules are catalyzed by large enzymatic complexes encoded by biosynthetic gene clusters (BGCs), which are often cryptic in that we do not understand what molecules they encode. Studying BGCs is challenging because (1) native fungi and bacteria can be difficult to culture in the lab and (2) expression in model organisms can result in seemingly inactive clusters. To address this challenge, we use cell-free protein synthesis to produce several gut bacterial Non-ribosomal peptide synthetases (NRPS) recently identified by a metagenomics analysis of the gut microbiome, making the first step toward prototyping some of these

BGCs *in vitro*. We express each open reading frame from the different BGCs individually by cell-free protein synthesis (CFPS) in *Escherichia coli* lysates. We find that at low temperatures we are able to increase the percent of full-length protein produced by CFPS. Interestingly, we observe multiple discrete truncation products in expression of our NRPSs. We anticipate that our efforts will inform future endeavors in expressing large protein by CFPS as well as pave the way toward prototyping these large natural product clusters.

8.2 Introduction

Natural products represent a diverse class of chemical compounds with a broad range of chemical structures and bioactivities[141,259-262]. Many of these compounds have become very successful pharmaceuticals. The natural product classes of polyketides (PKs) and non-ribosomal peptides (NRPs) constitute nearly all antibiotics. With ~70% of hospital-acquired bacterial infections are resistant to at least one antibiotic and ~40% resistant to three (CDC, USA), the discovery of new antibiotics and drugs is essential. However, less than 10% of the world's biodiversity has been evaluated for bioactivity. With some 50% of all drugs derived or inspired from natural sources, natural products represent an incredible amount of untapped potential[263,264]. If one in one hundred of the tens of millions of species estimated to be present in the world produced at least one novel natural product, there would be a huge reservoir of undiscovered molecules with potential medicinal value[262].

Historically, natural products have been found through the identification of their activities in nature. Through bioactivity-guided fractionation of organism extracts, the molecule of interest was isolated and characterized[265]. This traditional method is slow, often taking many years, and has failed to deliver many novel compounds. The major fault is that most secondary metabolites are not produced in laboratory fermentation conditions. On average, ~20 natural product biosynthesis clusters have been found in each newly sequenced bacterial genome, but only 1-2 products on average identified yielding the clusters cryptic[262]. Many new computational approaches being developed to identify and decrypt novel bacterial gene clusters[266]. For example, algorithms such as ClusterFinder and antiSMASH are two approaches to

identify natural products from gene clusters and vice versa[267]. These techniques have been used to identify gene clusters from a variety of phylogeny.

Recently, the human microbiome has been discovered to play a key role in modulating the human neuro-immuno-endocrine system[268-270]. This discovery has led to excitement in engineering and targeting the human microbiome in novel therapeutic strategies. A further look in to the composition of human microbiomes has shown that the microbial diversity and microbe-microbe interactions are responsible in part in digestion, homeostasis, and disease. Systematic analysis of microbial genomes of the human microbiome has led to the discovery of conserved biosynthetic gene clusters encoding potentially pharmaceutically relevant PKs and NRPs, some of which are exclusive to the human gut[271]. Producing these molecules in model organisms would greatly enhance the discovery and synthesis of novel drug candidates.

While the number of tools for natural product identification is growing, many of the responsible enzymes identified are large (>100kD). Large proteins are difficult to synthesize in E. coli and can be difficult in other model organisms. This slows down the ability to build natural product pathways. A more modern and thorough method of identifying, producing, and modifying natural products needs to be found.

Thousands of bacterial strains constitute the human microbiome with bacterial compositions differing between people and body sites within the same person. The small-molecule products made by these bacteria are catalyzed by enzymes encoded by biosynthetic gene clusters (BGCs). Recently, a metagenomic search of the human microbiota discovered a large family of short, relatively conserved nonribosomal peptide (NRP) BGCs exclusive to gut bacteria. The NRP products of these genetic elements represent gaps in our knowledge of what the microbiota can produce and how they mediate interspecies interactions. To study these clusters, scientists often resort to expressing them in model organisms because gut bacteria can be difficult to culture in the lab. However, expression in model organisms can result in no product formation due to seemingly inactive enzymes.

Through several collaborations we have begun to study how cell-free systems can be used to express and characterize NRPSs and natural products. In this work, we look at four classes of natural product enzymes: (1) relatively conserved, short NRPSs from gut bacteria (**Figure 8.1**), (2) short NRPSs and their modules from fungi, (3) the enterobactin biosynthetic gene cluster, and (4) engineered polyketide

synthase (PKS) molecules. We provide an in-depth analysis of and the issues associated with cell-free expression of these large (>100kDa) proteins as well as the metabolite production from a subset of these enzymes. As we learn more about how to express and characterize large natural product gene clusters our hope is use cell-free systems to better understand natural product biosynthesis, discover new natural products, and engineer existing natural products with new functions.
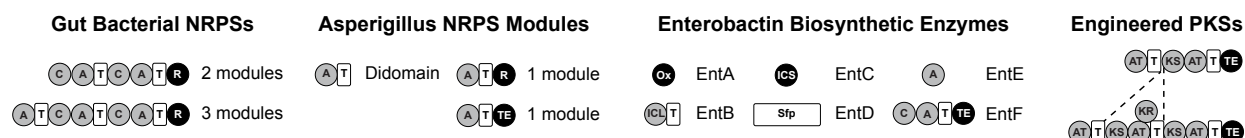
Figure 8.1. Overview of the classes of natural product enzymes surveyed. *We looked at short, conserved gut bacterial NRPS gene clusters, NRPS modules from Asperigillus species, the enterobactin biosynthetic gene cluster from* E. coli, *and engineered PKSs.*

8.3   Materials and Methods

8.3.1   Bacterial strains and plasmids

E. coli BL21(DE3) (NEB) was used for preparation of cell extracts. For *in vitro* expression of proteins, the pJL1 vector with a kanamycin (50 µg ml^{-1}) resistance selection marker was used. Detailed protocols are described in Chapter 4.

8.3.2   Cell Extract Preparation

E. coli BL21(DE3) cells were grown, harvested, lysed, and prepared using protocols in Chapter 4.

8.3.3   CFPS Reactions

CFPS reactions were performed to express BGC enzymes using methods described in Chapter 4 as well as in other pubications.[152,153] A 50 µL CFPS reaction in a 2.0 mL microcentrifuge tube was prepared and incubated at various temperatures noted throughout the text.

8.3.4   *In vitro* small molecule synthesis reactions

CFPS reactions were mixed with HEPES at 50 mM, substrates (e.g., DHB, amino acids, etc.), CoA, and NAD(P)H, when necessary noted throughout the text.

8.3.5   Metabolite Quantification

Metabolite extracts were analyzed by LC-MS using a Thermo Q Exactive mass spectrometer in line with an Agilent 1200 HPLC as described previously.[272]

8.4   Results

Natural products represent a class of molecules made from large multi-modular enzymes with complex and beneficial functions. A majority of our analysis has been focused on the expression of gut bacterial NRPS gene clusters (BGCs) (**Figure 8.2**).[271] In addition, we comment on the expression and characterization of short NRPS modules from fungi, the enterobactin biosynthetic cluster, and engineered polyketide synthase molecules. This study surveys the potential and limitations of using cell-free systems to study and produce natural products.
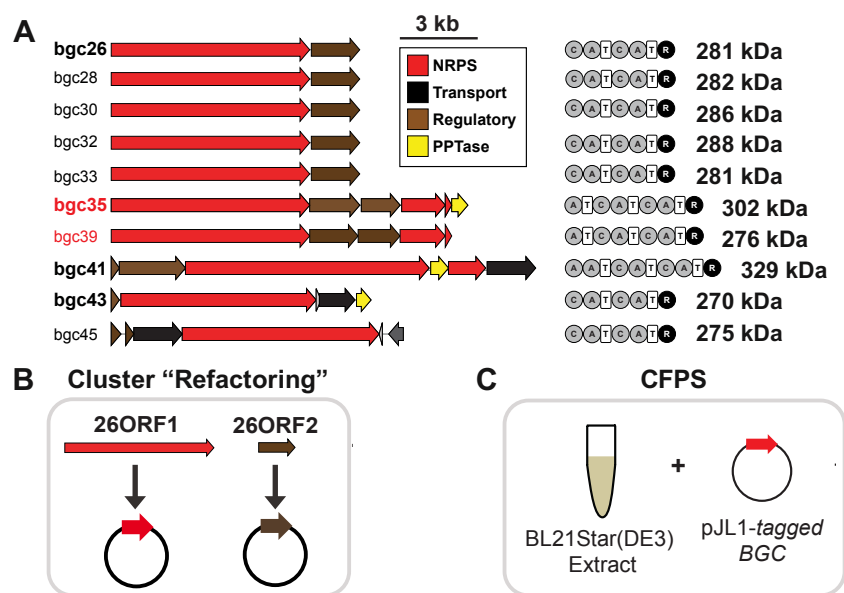
Figure 8.2. Schematic of gut BGCs and the cell-free approach. (A) ten gut BGCs were selected[271] ranging from 270 to 329 kDa in size of the largest protein in the cluster. (B) Cluster "refactoring" was performed by cloning out each open reading frame into the pJL1 backbone. (C) CFPS was used to express each protein.

8.4.1   Gut bacterial gene clusters

We selected two gut BGCs, BGC26 and BGC41, to learn how to express large enzymes in cell-free systems. We first tried to express these two BGCs, 6 proteins total, by CFPS in BL21 Star(DE3) *E. coli* extracts. We incorporated radioactively labelled leucine to observe protein products by autoradiogram

(**Figure 8.3**). This approach did not produce full-length proteins for each and every protein above 100 kDa. However, the smaller ancillary proteins were produced at full-length; BGC41-ORF4 is a transmembrane protein which smear on denaturing gels. To alleviate this issue, we tested CFPS of these proteins at five temperatures: 16, RT, 30, 37, and 42 ºC. Our key finding is that at low temperatures large proteins are easier to express at full-length (**Figure 8.3**). This aligns with *in vivo* approaches for troubleshooting protein expression as lowering the temperature of expression of difficult-to-express proteins is common in living *E. coli*. However, the several, distinct bands on the autoradiograms suggest that the proteins, while some are full-length, are being truncated.
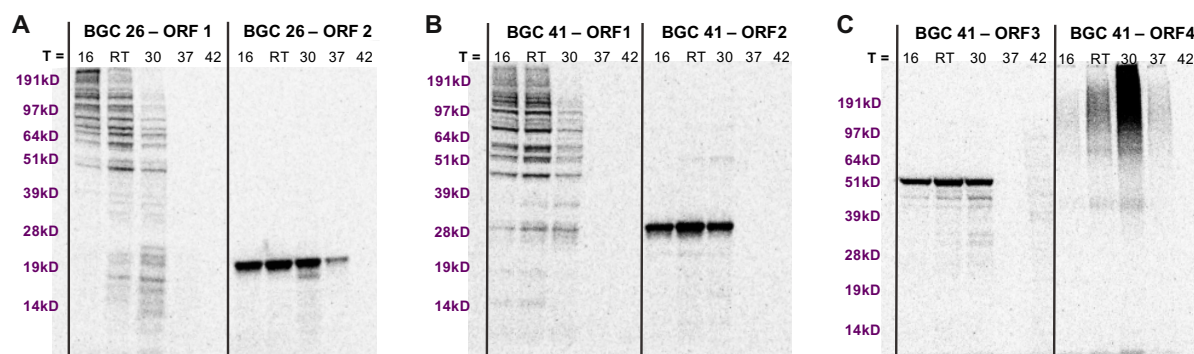
Figure 8.3. Low temperatures increase full-length protein production by CFPS. *CFPS with 14C leucine incorporation was performed at 16, RT, 30, 37, and 42 ºC for 20 h. Autoradiograms were run to observe protein produced for (A) BGC26-ORF1, BGC26-ORF2, (B) BGC41-ORF1, BGC41-ORF2, (C) BGC41-ORF3, and BGC41-ORF4.*

We try a variety of methods to determine where the truncated bands originate from looking at the DNA, RNA, and protein levels. First, we ran an agarose gel to observe DNA after a CFPS reaction of the first open reading frame of BGC26, 30, 32, and 45 (**Figure 8.4A**). We run DNA just after the CFPS reaction and cut with the NdeI restriction enzyme to view linear DNA. The gel shows that DNA is not significantly degraded or cleaved in our CFPS reactions. Second, we ran an RNA TBE-Urea gel of RNA extracted from our CFPS reactions of BGC26 and BGC45 at 0, 20, 40, 60, 120, and 240 minutes (**Figure 8.4B**). While messy this gel suggests that our RNA that our RNA is not being degraded over time. Third, we ran ribosome binding site computational predictions for BGC26, 30, 32, and 45 and find the with version 2.0 of the Salis RBS calculator that there is close to no potential internal ribosome binding sites (IRESs) within the BGC sequences (**Figure 8.4C**). Thus, IRESs are likely not contributing to truncated bands. Fourth, we added

protease inhibitors to our CFPS reactions of BGC45, 26, and 41 to see if protein cleavage could be the reason for our truncated bands. However, we observe that proteases targeted by our protease inhibitor cocktails do not seem to be the reason for our truncated bands (**Figure 8.4D**). In addition, we tagged BGC26, 30, 32, and 45 with N-terminal FLAG tags and C-terminal HIS tags running them in CFPS. Based on fluorescent probing of these modification we see that the truncations are all over the protein (**Figure 8.5**). Protease cleavage could be further looked into as a source of these bands. DNA, RNA, and protein degradation are not likely reasons for the peptide banding pattern.
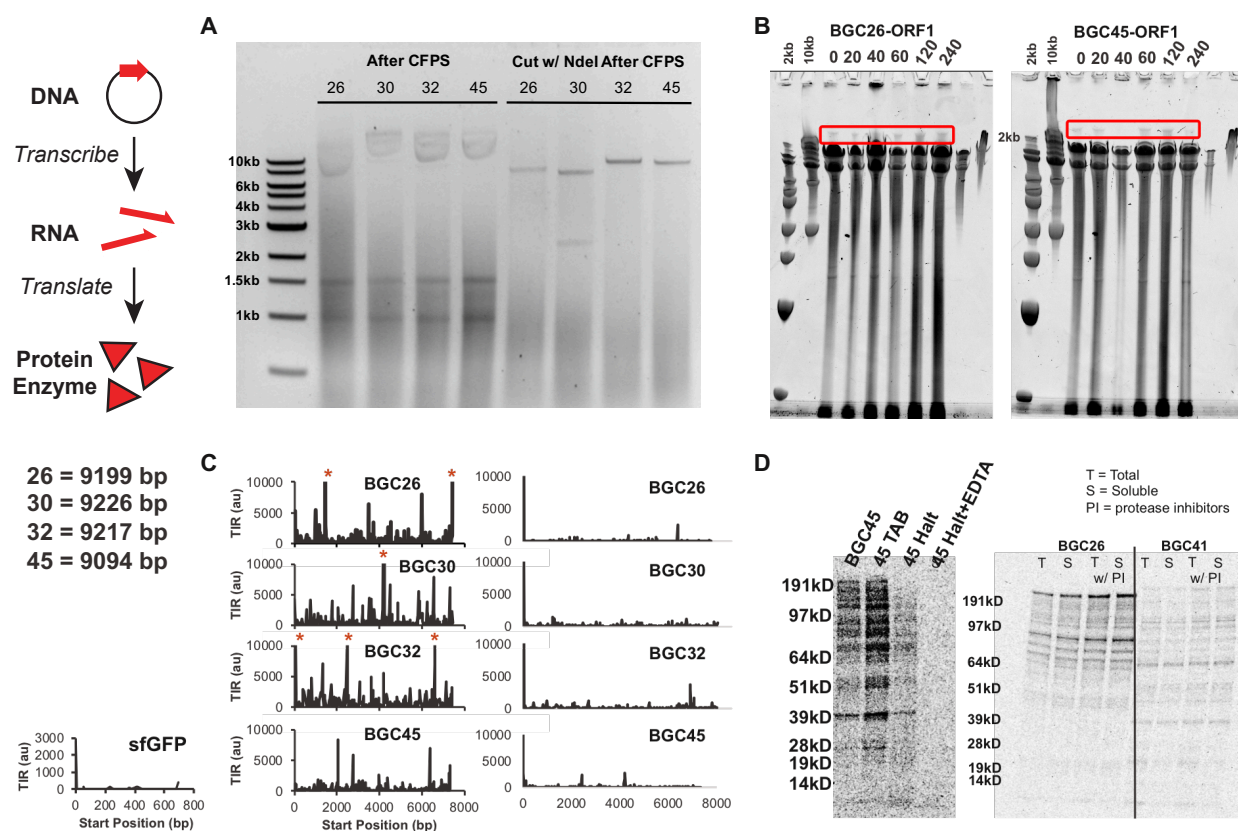
Figure 8.4. Four methods to decipher truncation products by CFPS. *(A) NRPS-encoded DNA was run on a 1% agarose gel after a 20 h CFPS reaction. The reaction was directly added to the gel and separately was cut with NdeI and added to the gel. (B) mRNA from CFPS reactions was extracted at 0, 20, 40, 60, 120, and 240 min of reaction time and run on an RNA TBE-UREA gel with a 2 kB and 10 kB ladder. (C) Internal ribosome binding sites (IRESs) were computationally determined with the Salis ribosome binding site calculator. Version 1.0 is on the left and version 2.0 is on the right. (D) Roche protease inhibitor tablets, Halt protease inhibitor cocktail, and Halt plus EDTA were added at the beginning of a 20 h CFPS reaction. These reactions were run on an SDS-PAGE gel autoradiogram.*
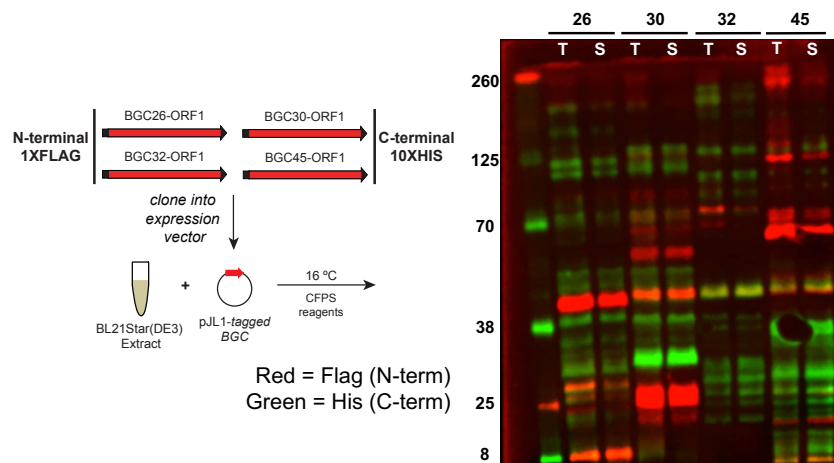
Figure 8.5. N-terminal and C-terminal tags provide observation of truncations. BGC26, BGC30, BGC32, and BGC45 were tagged with 1xFLAG on the N-terminus and 10xHIS on the C-terminus. Tagged proteins were expressed by CFPS for 20 h at 16 ºC then run on an SDS-PAGE gel. Western blot detection was observed with anti-HIS and anti-FLAG antibodies.

Moving forward, we are able to use cell-free systems to make several full-length NRPSs, some of which have unknown products (**Figure 8.6A-B**). While the truncated banding pattern is present, the enzymes themselves may still be intact just cleaved in loops. We tried to express this in Streptomyces lividans cell-free systems with the hypothesis that Streptomyces has several native natural product pathways and might be able to express natural product enzymes better than E. coli. We find that our Streptomyces cell-free platform cannot produce these gut BGCs (**Figure 8.6C**). However, from **Figure 8.6A** we decided to move forward to using the BGCs to make metabolites.
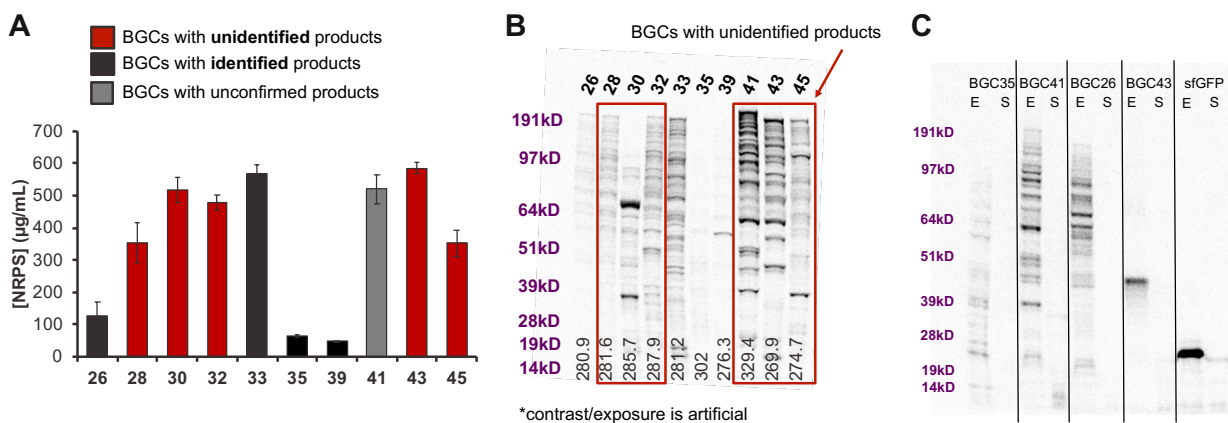
Figure 8.6. Several full-length non-ribosomal peptide synthetases can be made by CFPS. CFPS reactions of all ten BGCs were run. (A) These reactions were quantified by radioactivity. (B) An SDS-PAGE gel and autoradiogram were run and analyzed. (C) CFPS reactions for BGC35, BGC41, BGC26, and BGC43 were run and autoradiograms are shown.

While we are unsure which metabolites should be made from these BGCs, we know that several identified gut BGC NRPSs make dipeptides.[273] Thus, we took BGC26 and tried to look for dipeptides from

its activity. We see that we get unique chromatogram peaks in BGC26 samples corresponding to $C_{22}H_{45}N_2O_3$–consistent with a dipeptide expected from BGC26 NRPS (**Figure 8.7**). However, the peaks are very low intensity, only ~10-fold above noise level, and no MS2 scan was acquired for the ion. From here we have moved forward to further analyze these BGCs and their metabolites using cell-free systems.
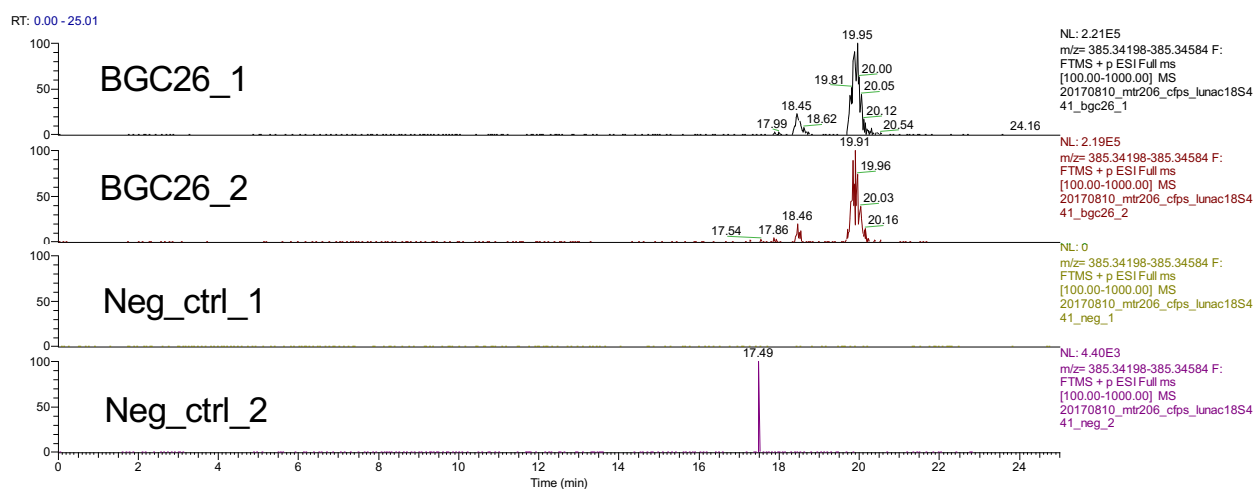
Figure 8.7. Metabolite production from BGC26. *2 samples of BGC26 biosynthesis reactions and 2 negative controls (No DHB added) were run for 20 h. LC-MS Chromatograms for each are displayed.*

8.4.2   Short NRPS modules from fungi

NRPSs come from many domains of life including fungi. We decided to explore the NRPSs from Aspergillus species, going beyond bacterial natural products. First, we looked at expressing three AT didomains (~75 kDa) (**Figure 8.8A**). These modules have some new fatty amino acid loading activities hypothesized but needs to be proven by direct observation. We took Asperigillus DNA encoding these modules and cloned them into pJL1 for CFPS. We were able to get robust expression of these three proteins. Furthermore, we tried to look at the activity of AT1. We detected AT1 in the apo, holo, loaded forms (**Figure 8.8B**). However, from analysis we were not able to detect properly loaded AT1 rather we saw what seemed to be an acetyl and a C18 loaded. Perhaps, these were loaded from the diverse small molecule background in the cell-free reactions. We had no success trying to troubleshoot why the loaded forms are not detected.
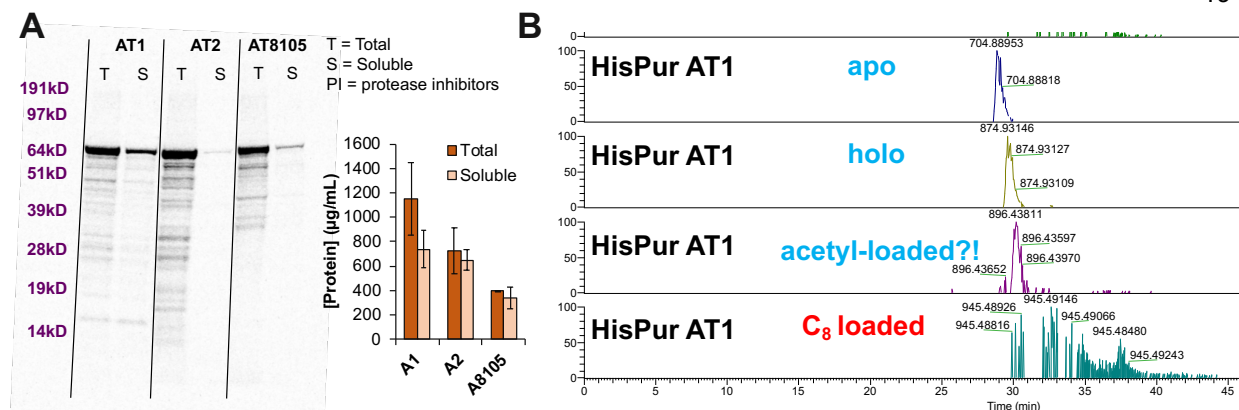
Figure 8.8. Characterization of AT didomains from Asperigillus DNA. *Three CFPS reactions were performed for 20 h of AT1, AT2, and AT8105. (A) These reactions were quantified by radioactivity and an autoradiogram shows protein products from each reaction. (B) CFPS reactions of AT1 were then mixed with HEPES buffer, Sfp, ATP, and amino acids. These biosynthesis reactions were then HIS purified and run on LC-MS to determine loading onto the T domains.*

Next, we looked at five small NRPS single modules that naturally exist in different Asperigillus species four that produce unknown metabolites and one that produces aspulvinone. We selected these because they are short and should be relatively easy to express by CFPS and they have potentially interesting biochemistries (**Figure 8.9A**). We expressed each of these by CFPS and observed testable titers of protein from each, though we see the same banding patterns we saw with our gut BGCs (**Figure 8.9B-C**). Using a fluorescently labeled CoA assay, we are able to detect faint activity from the A-T-R1 and the aspulvinone A-T-TE module (**Figure 8.9D**). However, no detectable product was observed by LC-MS.
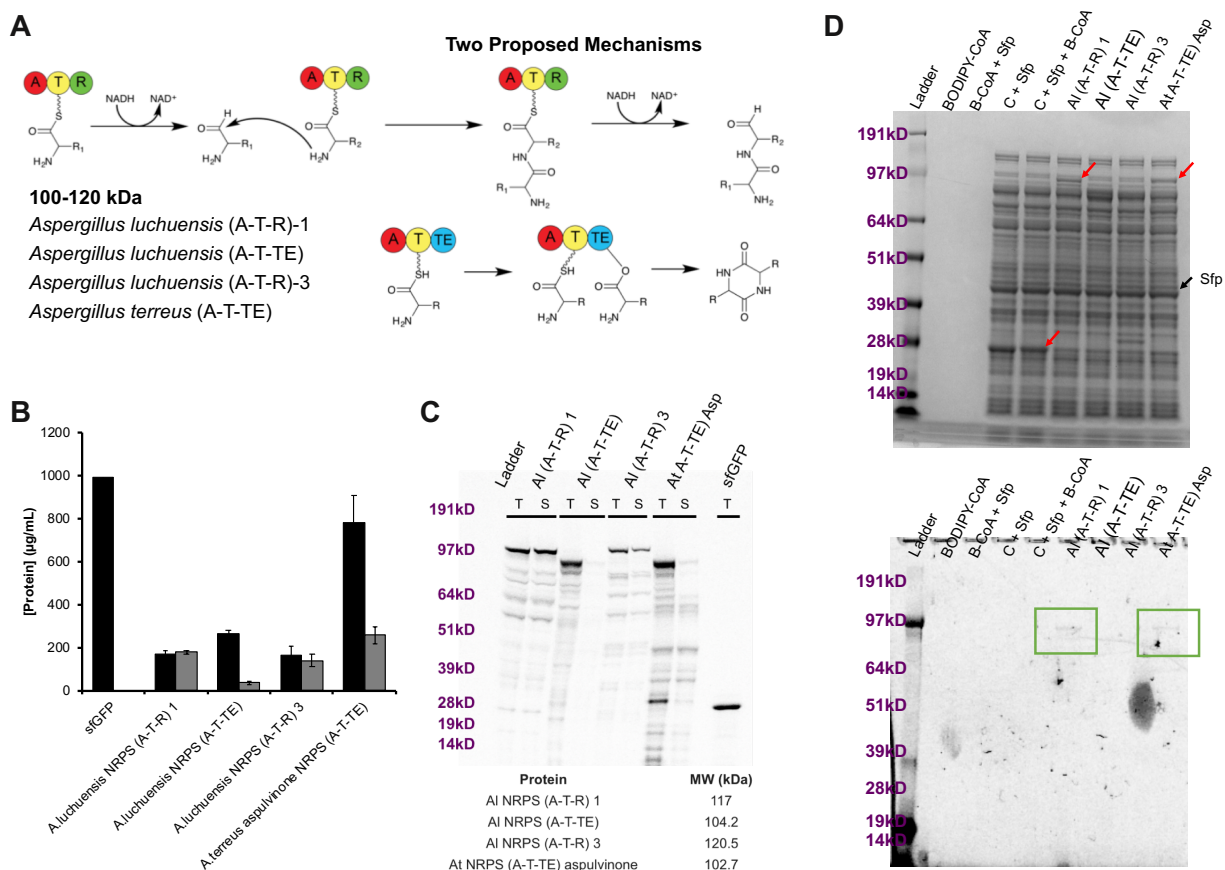
Figure 8.9. Expression of Asperigillus one-module NRPSs. *(A) Schematic representation of two proposed mechanisms for A-T-R and A-T-TE modules. (B) Radioactive analysis of CFPS reactions of A. luchuensis NRPS ATR-1, A. luchuensis NRPS ATTE, A. luchuensis NRPS ATR-3, and A. terreus aspulvinone NRPS ATTE. Black bars are total protein amounts and gray bars are soluble protein amounts. (C) Autoradiogram analysis of each NRPS module. (D) Bodipy-CoA and Sfp were added to CFPS reactions of A. luchuensis NRPS ATR-1, A. luchuensis NRPS ATTE, A. luchuensis NRPS ATR-3, and A. terreus aspulvinone NRPS ATTE to detect loading of NRPS modules. These reactions were run on SDS-PAGE to determine loading of each NRPS module.*

8.4.3   Enterobactin biosynthetic gene cluster

We studied cell-free production of enterobactin, a siderophore naturally produced by E. coli in iron limited environments but not in cell-free extracts. This pathway involves six proteins EntA-F to create the cyclic enterobactin (**Figure 8.10A**). We first expressed each of the six enzymes via CFPS and were able to produce near gram per liter quantities for many of these proteins (**Figure 8.10B**). From here we decided to validate the activity of the last step of the pathway requiring EntB, EntE, and EntF. We produced each of these separately by CFPS, mixed them under different physiochemical conditions, and supplied

dihydroxybenzoate (DHB) to initiate the reaction. After 20 h of incubation at 30 ºC we observed enterobactin and DHB present by LC-MS. We see that, while the on intensity is maybe 10-fold above the limit of detection, the data suggests that enterobactin can be made in cell-free systems from in vitro-expressed EntB, EntE, and EntF (**Figure 8.10C-D**). The yields seem to be in the range of ~0.1 µg/L.
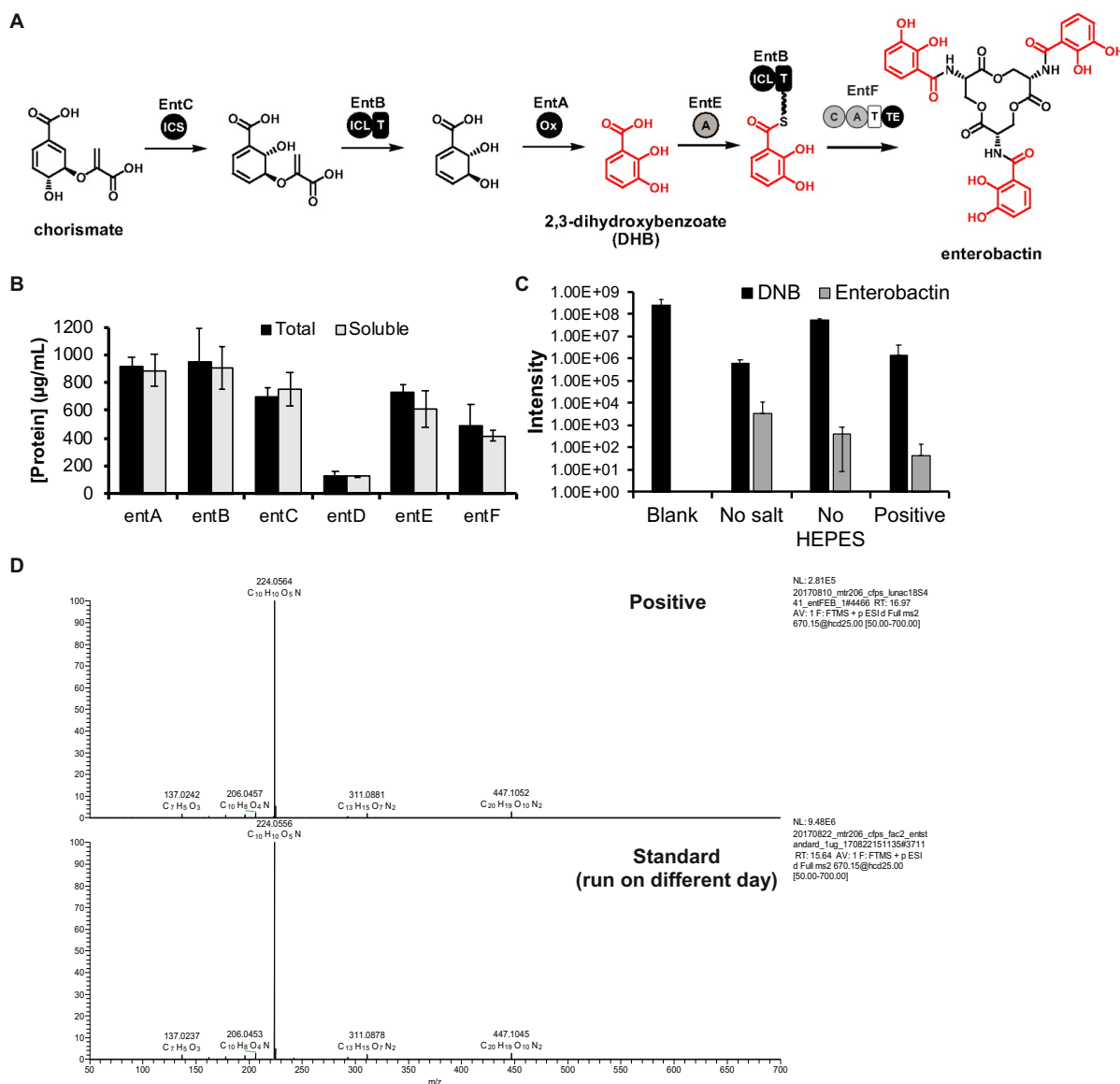
Figure 8.10. Characterization of Enterobactin biosynthetic pathway. *(A) Schematic representation of enterobactin biosynthetic pathway. (B) CFPS analysis of total and soluble yields of each enzyme in the enterobactin biosynthetic gene cluster. (C) CFPS reactions of EntE, EntB, and EntF were mixed with DHB, HEPES buffer and Mg^{2+}, NH_4^+, and K^+ salts (Positive). Salts and HEPES were also individually left out. A Blank was including containing only E. coli extract. After 20 h reactions, samples were analyzed for DHB and enterobactin by LC-MS. (D) MS2 spectrum for a reaction with EntE, EntB, and EntF were mixed with DHB, HEPES buffer and Mg^{2+}, NH_4^+, and K^+ salts (Positive) as well as a commercial standard are shown.*

8.4.4   Engineered polyketide synthases

Engineering polyketide synthases (PKSs) would allow us to discover design principles for PKSs so that they can be readily constructed for producing many different biofuels and bioproducts. We use cell-free systems to test chimeric PKSs. Specifically, we look at PKSs from DEBS, LIP, and FLU biosynthetic pathways. These are the largest proteins we have tried to express in cell-free reactions (~400 kDa). We expressed them by CFPS and measure protein production for each (**Figure 8.11**). These samples are sent to LBL and EMSL for enzymatic assays and analysis.
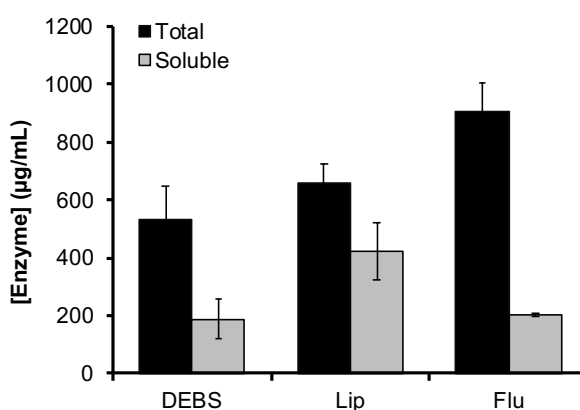
Figure 8.11. CFPS of engineered PKSs. CFPS reactions were run at 16 ºC for 20 h of DEBSPKS-GFP, LipPKS-GFP, and FluPKS-GFP fusion proteins. Radioactivity is analyzed.

8.5 Summary

In conclusion, temperature is a key lever in expressing large proteins and could help in realizing the potential of cell-free protein synthesis systems to study natural product biosynthesis. Truncated banding patterns are present in large amounts when expressing large natural product enzymes in cell-free systems and are likely not due to DNA, RNA, or protein degradation. Further investigation of these banding patterns is necessary. Another interesting result from these studies is that while many are able to be expressed at relatively high soluble quantities, not all of the soluble protein product is active or that activity is low. Thus, we have had to add much more enzyme than seemingly required (>1 μM quantities). Overall, we believe that cell-free systems have great potential in studying and discovering natural products and the enzymes that produce them.

8.6 Acknowledgments

This work would not be possible without the outstanding collaborations that have catalyzed these works. We would like to thank Michael Fischbach and Chunjun Guo at UCSF for providing DNA for the BGCs used in this study. Special thanks go out to Matt Robey, Matt Henke, and Neil Kelleher for collaboration in studying gut BGC metabolite production, the fungal NRPS modules, and the enterobactin work. We would also like to thank The Keasling lab for the collaboration of the engineered PKSs.

8.7 Conclusions

This chapter described several experiments to make gut bacterial non-ribosomal peptide synthetases (NRPSs), fungal NRPSs, an NRPS derived from *E. coli*, and engineered polyketide synthases (PKSs). We had success in producing these proteins by CFPS and some success in having them make their metabolite products. We learned the temperature is an important lever in attaining full-length expression of large proteins and that large proteins have a protein gel banding patterns derived from some form of protein truncation. We also find that adding higher concentrations of enzyme than seemingly necessary is important to observe enzyme activity.

Chapter 9: Achievements and future directions

The beginning of knowledge is the discovery of something we do not understand.

- Frank Herbert

9.1 Achievements of my research

The overarching goal of my dissertation research has been to develop powerful technologies to rationally and rapidly manipulate biological systems to produce chemicals. For decades, scientists and engineers have turned to biological systems to help meet societal needs in energy, medicine, materials, and more—especially when chemical synthesis is untenable (e.g., antimalarial drugs). Often, biologically-produced small molecules are insufficient for production at the optimal titer, rate, or yield because natural sources are difficult to optimize and are simply not scalable (e.g., plants grow slowly). Thus, engineers seek to design enzymatic reaction schemes in model microorganisms to meet manufacturing criteria. Success in these endeavors depends upon identifying sets of enzymes that can convert readily available molecules (e.g., glucose) to high-value products (e.g., medicines), with each enzyme performing one of a series of chemical modifications. Unfortunately, this is difficult because design-build-test (DBT) cycles—iterations of re-engineering organisms to test new sets of enzymes—are detrimentally slow due to the constraints of cell growth. As a result, a typical project today might only explore dozens of variants of an enzymatic reaction pathway. This is often insufficient to identify a commercially relevant solution because selecting productive enzymes using existing single-enzyme kinetic data has limited applicability in multi-enzyme pathways and consequently requires more DBT iterations. *With nearly half of approved small molecule drugs being derived from natural products and nearly all chemicals being produced from petroleum, it is essential that*

we speed up the biochemical discovery process. My research sought to address these problems by re-conceptualizing the way we engineer and unearth enzymatic pathways using cell-free systems.

In my time at Northwestern, I built and developed a novel cell-free framework utilizing *in vitro* protein synthesis for enzymatic pathway prototyping and discovery. The foundational principle is that we can construct discrete enzymatic pathways through modular assembly of cell lysates containing enzymes produced by cell-free protein synthesis rather than by living organisms. This provides an unprecedented capability to test hundreds of thousands of pathways by avoiding inherent limitations of cell growth and thus diminishing the reliance on single-enzyme kinetic data. A key conceptual innovation is that the DBT unit can be cellular lysates rather than genetic constructs, allowing us to perform DBT iterations without the need to re-engineer organisms. I established this approach for production of *n*-butanol, 3-hydroxybutyrate, and acetone, industrially-relevant chemicals, showing that *E. coli*-cell lysates support active enzymatic pathways and functional metabolism (Chapter 3);[173] the fact that this works is still surprising to many in the field. I have presented this work at multiple conferences and was awarded 2nd prize at the international synthetic biology conference, SEED, in 2016. To further develop the framework's flexibility, I optimized every component of the cell-free chemical environment to substantially enhance our ability to build and test cell-free enzymatic pathways (Chapter 6).[170] My framework has attracted both academic and industrial collaborations including one with LanzaTech, a world leader in converting carbon waste to valuable chemicals using microbes. With them, I demonstrated that cell-free experimentation can indeed inform cellular biochemical production, a grand challenge in the field (Chapter 7). To this end, I leveraged liquid-handling robotics to access combinatorial design space exceeding typical pipelines pursued in cells (Chapter 5) and created a new metric capturing *in vitro* titers, rates, and enzyme expression to inform *in vivo* pathway construction (Chapter 7).

Another tenet of my research has focused on biomolecule discovery. Medicines are often derived from fungal and bacterial natural products. Many of these molecules are catalyzed by large enzymatic complexes encoded by biosynthetic gene clusters (BGCs). In my last years at Northwestern, I addressed challenges associated with studying BGCs by expanding my cell-free framework to prototype BGCs from gut bacteria, fungi*, and E. coli* to study biosynthesis of known natural products and, importantly, attempt to

discover new ones (Chapter 8). It's importance to the field is already being recognized, evident by a poster award I received at the 2017 Engineering Biology Research Consortium meeting.

There are many highlights of my PhD career at Northwestern. My work is the first development of cell-free protein synthesis for enzymatic pathway construction and has spawned a growing interest in cell-free metabolic systems. Several companies and labs have already contacted me to collaborate and use my framework. In addition, my work has changed the trajectory of my subgroup within the Jewett Lab. I was even recognized with Northwestern's Chemical and Biological Engineering Department's 2017 Distinguished Graduate Researcher Award for my dissertation work. I was also invited to be a U.S. delegate to the 67th Lindau Nobel Laureates Meeting. In addition, I was finalist for the EURAXESS North American Science Slam where performed an interactive presentation of my research to a general audience. I anticipate that my work in tandem with high-end metabolomics will offer a high degree of flexibility to model the kinetics and stability of individual enzymes, measure metabolite fluxes in multistep pathways, and experimentally isolate many other parameters confounded in living organisms. Just as rapid increases in computing described by Moore's law have profoundly impacted the electronics industry, this work represents the next advancement in engineering that should have an analogous impact on synthetic biology.

The important thing is not to stop questioning. Curiosity has its own reason for existing.

- Albert Einstein

9.2 Future Directions

Cell-free systems have a bright future ahead. This is seen by the growing interest from funding and government agencies, growth that has all happened within the last five years. I predict that cell-free systems for biosynthetic pathway prototyping and biomanufacturing will also grow. Specifically, I see future

directions of my work in, but not limited to, three main areas: (1) mimicking cellular physiochemical conditions in the cell-free environment, (2) enzyme discovery, and (3) modified metabolites.

In Chapter 6 I described the in-depth look into controlling cell-free metabolism by changing the physiochemical environment in which the reactions are run. Learning how to control these for optimal biosynthetic pathway operation is still an ongoing area of research. Specifically, we still need to learn how to control pH and modulate it during a cell-free reaction. Balancing cofactors and their redox state is poorly understood in our systems and adapting regulators that are used in cells as well as other cell-free systems[103] will be essential moving forward. Using and removing native metabolism will also be important in controlling the biosynthesis of small molecules. Each of these will contribute to the repertoire of tools that will allow us to mimic metabolic states of cells, different periods of cellular growth, and a diverse array of bacteria. We will then be able develop a robust set of modelling tools to study cell-free metabolism and cell-free-to-cell correlations. Taken together these steps will lead to a much more powerful cell-free toolbox for cellular prototyping.

In Chapter 3 we discovered that several putative bifunctional enzymes indeed had bifunctionality, but we also deciphered that one had mono-functionality. This, combined with the study of natural product enzymes in Chapter 8, really highlights the potential of cell-free systems to be used for enzyme discovery. There is an exponentially growing number of genomes from new organisms being published, and with each one comes an even larger number of putative enzymes to test for novel and more-favorable chemistries. This is where I believe cell-free systems will shine. A better of understanding of truncations in cell-free systems is needed. Furthermore, development of high-throughput enzyme activity assays and sensors for enzyme activities are desperately needed. In study new enzymes for novel functions this will lead to new chemistries and new small molecules that can be made.

The discovery of new enzymes and chemistries will allow for the modification of small molecules metabolites that already exist. For example, enzymes that hydroxyl groups to chemicals at specific locations can be screened in cell-free for their promiscuity to hydroxylate new chemicals leading to modified metabolites. Mixing and matching these tailoring enzymes—like those that already exist for antibiotics— could give rise to new antibiotics and antibiotic properties in molecules that are not antibiotic at all. Modified metabolites also exist in the form of glycosides and sugar alternatives—plant metabolites are highly

modified and their biosyntheses are relatively unknown. Cell-free systems should be used to study these chemistries and how they can be harnessed to create novel chemicals.

Future directions of the work presented in this dissertation are not limited to the areas of prototyping for cellular metabolic engineering, enzyme discovery, or novel chemistry, but these represent some of the most-promising next steps. It is with pleasure, and hope, that I pass this along to the next members of the Jewett lab and to the scientific community at large. Cell-free systems are changing the biotechnology paradigm for the better.

People that complete other people's vision are understated.

- Björk

References

1   Sheldon, R. A. The E Factor: fifteen years on. *Green Chemistry* **9**, 1273, doi:10.1039/b713736m (2007).

2   Werpy, T. P., G. ; Aden, A. ; Bozell, J. ; Holladay, J. ; White, J. ; Manheim, Amy ; Eliot, D. ; Lasure, L. ; Jones, S. Top Value Added Chemicals From Biomass. Volume 1 - Results of screening for potential candidates from sugars and synthesis gas. *DTIC Document* **1** (2004).

3   Bozell, J. J. & Petersen, G. R. Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's "Top 10" revisited. *Green Chemistry* **12**, 539, doi:10.1039/b922014c (2010).

4   Nielsen, J. *et al.* Engineering synergy in biotechnology. *Nature chemical biology* **10**, 319-322, doi:10.1038/nchembio.1519 (2014).

5   Keasling, J. D. Manufacturing molecules through metabolic engineering. *Science* **330**, 1355-1358, doi:10.1126/science.1193990 (2010).

6   Lee, J. W., Kim, H. U., Choi, S., Yi, J. & Lee, S. Y. Microbial production of building block chemicals and polymers. *Current opinion in biotechnology* **22**, 758-767, doi:10.1016/j.copbio.2011.02.011 (2011).

7   Curran, K. A. & Alper, H. S. Expanding the chemical palate of cells by combining systems biology and metabolic engineering. *Metabolic engineering* **14**, 289-297, doi:10.1016/j.ymben.2012.04.006 (2012).

8   Nielsen, J., Larsson, C., van Maris, A. & Pronk, J. Metabolic engineering of yeast for production of fuels and chemicals. *Current opinion in biotechnology* **24**, 398-404, doi:10.1016/j.copbio.2013.03.023 (2013).

9   Galanie, S., Siddiqui, M. S. & Smolke, C. D. Molecular tools for chemical biotechnology. *Current opinion in biotechnology* **24**, 1000-1009, doi:10.1016/j.copbio.2013.03.001 (2013).

10   Keasling, J. D. Synthetic biology and the development of tools for metabolic engineering. *Metabolic engineering* **14**, 189-195, doi:10.1016/j.ymben.2012.01.004 (2012).

11   Woolston, B. M., Edgar, S. & Stephanopoulos, G. in *Annual Review of Chemical and Biomolecular Engineering, Vol 4* Vol. 4 *Annual Review of Chemical and Biomolecular Engineering* (ed J. M. Prausnitz) 259-288 (Annual Reviews, 2013).

12   Clomburg, J. M. & Gonzalez, R. Biofuel production in Escherichia coli: the role of metabolic engineering and synthetic biology. *Appl. Microbiol. Biotechnol.* **86**, 419-434, doi:10.1007/s00253-010-2446-1 (2010).

13   Kwok, R. Five hard truths for synthetic biology. *Nature* **463**, 288-290, doi:10.1038/463288a (2010).

14   Yadav, V. G., De Mey, M., Giaw Lim, C., Kumaran Ajikumar, P. & Stephanopoulos, G. The future of metabolic engineering and synthetic biology: Towards a systematic practice. *Metabolic engineering* **14**, 233-241, doi:10.1016/j.ymben.2012.02.001 (2012).

15   Nakamura, C. E. & Whited, G. M. Metabolic engineering for the microbial production of 1,3-propanediol. *Current opinion in biotechnology* **14**, 454-459, doi:10.1016/j.copbio.2003.08.005 (2003).

16   Renniger, N. M., D. Fuel compositions comprising farnesane and farnesane derivatives and method of making and using same. U.S. Patent No. 7399323. U.S. Patent No. 7399323 (2008).

17   Hodgman, C. E. & Jewett, M. C. Cell-free synthetic biology: thinking outside the cell. *Metabolic engineering* **14**, 261-269, doi:10.1016/j.ymben.2011.09.002 (2012).

18   Guterl, J.-K. & Sieber, V. Biosynthesis "debugged": Novel bioproduction strategies. *Engineering in Life Sciences* **13**, 4-18, doi:10.1002/elsc.201100231 (2013).

19   Swartz, J. R. Transforming biochemical engineering with cell-free biology. *AIChE Journal* **58**, 5-13 (2011).

20   Zhang, Y. H. P. What is vital (and not vital) to advance economically-competitive biofuels production. *Process Biochemistry* **46**, 2091-2110, doi:10.1016/j.procbio.2011.08.005 (2011).

21   Ardao, I. & Zeng, A.-P. In silico evaluation of a complex multi-enzymatic system using one-pot and modular approaches: Application to the high-yield production of hydrogen from a synthetic metabolic pathway. *Chemical Engineering Science* **87**, 183-193, doi:10.1016/j.ces.2012.10.005 (2013).

22   Rollin, J. A., Tam, T. K. & Zhang, Y. H. P. New biotechnology paradigm: cell-free biosystems for biomanufacturing. *Green Chemistry* **15**, 1708, doi:10.1039/c3gc40625c (2013).

23   Guterl, J. K. *et al.* Cell-free metabolic engineering: production of chemicals by minimized reaction cascades. *ChemSusChem* **5**, 2165-2172, doi:10.1002/cssc.201200365 (2012).

24   Murray, C. J. & Baliga, R. Cell-free translation of peptides and proteins:from high throughput screening to clinical production. *Curr. Opin. Chem. Biol.* **17**, 420-426, doi:http://dx.doi.org/10.1016/j.cbpa.2013.02.014 (2013).

25   Catherine, C., Lee, K.-H., Oh, S.-J. & Kim, D.-M. Cell-free platforms for flexible expression and screening of enzymes. *Biotechnol. Adv.* **31**, 797-803, doi:http://dx.doi.org/10.1016/j.biotechadv.2013.04.009 (2013).

26   Chappell, J., Jensen, K. & Freemont, P. S. Validation of an entirely *in vitro* approach for rapid prototyping of DNA regulatory elements for synthetic biology. *Nucleic Acids Res.* **41**, 3471-3481, doi:10.1093/nar/gkt052 (2013).

27   Shin, J. & Noireaux, V. An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and artificial cells. *ACS Synth. Biol.* **1**, 29-41, doi:10.1021/sb200016s (2012).

28   Jewett, M. C., Fritz, B. R., Timmerman, L. E. & Church, G. M. *In vitro* integration of ribosomal RNA synthesis, ribosome assembly, and translation. *Mol. Syst. Biol.* **9**, 678, doi:doi:10.1038/msb.2013.31 (2013).

29   Goerke, A. R. & Swartz, J. R. High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids. *Biotechnol. Bioeng.* **102**, 400-416, doi:10.1002/bit.22070 (2009).

30   Bundy, B. C. & Swartz, J. R. Site-specific incorporation of *p*-propargyloxyphenylalanine in a cell-free environment for direct protein−protein click conjugation. *Bioconjug. Chem.* **21**, 255-263, doi:10.1021/bc9002844 (2010).

31   Albayrak, C. & Swartz, J. R. Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation. *Nucleic Acids Res.* **41**, 5949-5963, doi:10.1093/nar/gkt226 (2013).

32   Hong, S. H. *et al.* Cell-free protein synthesis from a release factor 1 deficient *Escherichia coli* activates efficient and multiple site-specific nonstandard amino acid incorporation. *ACS Synth. Biol.*, online (2014).

33   Ugwumba, I. N. *et al.* Using a genetically encoded fluorescent amino acid as a site-specific probe to detect binding of low-molecular-weight compounds. *Assay Drug Dev. Technol.* **9**, 50-57 (2011).

34   Ugwumba, I. N. *et al.* Improving a natural enzyme activity through incorporation of unnatural amino acids. *J. Am. Chem. Soc.* **133**, 326-333, doi:10.1021/ja106416g (2010).

35   Mukai, T. *et al.* Genetic-code evolution for protein synthesis with non-natural amino acids. *Biochem. Biophys. Res. Commun.* **411**, 757-761, doi:10.1016/j.bbrc.2011.07.020 (2011).

36   Loscha, K. V. *et al.* Multiple-site labeling of proteins with unnatural amino acids. *Angew. Chem. Int. Ed. Engl.* **51**, 2243-2246, doi:10.1002/anie.201108275 (2012).

37   Swartz, J. R. Developing cell-free biology for industrial applications. *Journal of industrial microbiology & biotechnology* **33**, 476-485, doi:10.1007/s10295-006-0127-y (2006).

38   Rupp, S. Next-generation bioproduction systems: Cell-free conversion concepts for industrial biotechnology. *Engineering in Life Sciences* **13**, 19-25, doi:10.1002/elsc.201100237 (2013).

39   Billerbeck, S., Harle, J. & Panke, S. The good of two worlds: increasing complexity in cell-free systems. *Current opinion in biotechnology* **24**, 1037-1043, doi:10.1016/j.copbio.2013.03.007 (2013).

40   Kohler, R. E. The reception of Eduard Buchner's discovery of cell-free fermentation. *Journal of the History of Biology* **5**, 327-353, doi:10.1007/bf00346663 (1972).

41   Welch, P. & Scopes, R. K. Studies on cell-free metabolism: Ethanol production by a yeast glycolytic system reconstituted from purified enzymes. *Journal of Biotechnology* **2**, 257-273, doi:10.1016/0168-1656(85)90029-x (1985).

42   Nirenberg, M. W. & Matthaei, J. H. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences* **47**, 1588-1602, doi:10.1073/pnas.47.10.1588 (1961).

43   Swartz, J. R. Transforming Biochemical Engineering with Cell-Free Biology. *Aiche J* **58**, 5-13, doi:10.1002/aic.13701 (2012).

44   Zhu, F. *et al.* In vitro reconstitution of mevalonate pathway and targeted engineering of farnesene overproduction in Escherichia coli. *Biotechnology and bioengineering* **111**, 1396-1405, doi:10.1002/bit.25198 (2014).

45   Zaks, A. Industrial biocatalysis. *Current Opinion in Chemical Biology* **5**, 130-136, doi:10.1016/s1367-5931(00)00181-2 (2001).

46   Bogorad, I. W., Lin, T. S. & Liao, J. C. Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature* **502**, 693-697, doi:10.1038/nature12575 (2013).

47   Jensen, V. J. & Rugh, S. [33] Industrial-scale production and application of immobilized glucose isomerase. **136**, 356-370, doi:10.1016/s0076-6879(87)36035-5 (1987).

48   Bruggink, A., Roos, E. C. & de Vroom, E. Penicillin Acylase in the Industrial Production of β-Lactam Antibiotics. *Organic Process Research & Development* **2**, 128-133, doi:10.1021/op9700643 (1998).

49   Bujara, M. & Panke, S. *In silico* assessment of cell-free systems. *Biotechnology and bioengineering* **109**, 2620-2629, doi:10.1002/bit.24534 (2012).

50   Rosenthaler, L. Durch enzyme bewirkte asymmetrische synthesen. *Biochem. Z* **14**, 238-253 (1908).

51   Jewett, M. C., Calhoun, K. A., Voloshin, A., Wuu, J. J. & Swartz, J. R. An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol Syst Biol* **4**, 220, doi:10.1038/msb.2008.57 (2008).

52   Chen, X. *et al.* Statistical Experimental Design Guided Optimization of a One-Pot Biphasic Multienzyme Total Synthesis of Amorpha-4,11-diene. *PloS one* **8**, e79650, doi:10.1371/journal.pone.0079650 (2013).

53   Zhang, Y. H., Evans, B. R., Mielenz, J. R., Hopkins, R. C. & Adams, M. W. High-yield hydrogen production from starch and water by a synthetic enzymatic pathway. *PloS one* **2**, e456, doi:10.1371/journal.pone.0000456 (2007).

54   Ye, X. *et al.* Spontaneous high-yield production of hydrogen from cellulosic materials and water catalyzed by enzyme cocktails. *ChemSusChem* **2**, 149-152, doi:10.1002/cssc.200900017 (2009).

55   Martin del Campo, J. S. *et al.* High-yield production of dihydrogen from xylose by using a synthetic enzyme cascade in a cell-free system. *Angewandte Chemie* **52**, 4587-4590, doi:10.1002/anie.201300766 (2013).

56   Wang, Y., Huang, W., Sathitsuksanoh, N., Zhu, Z. & Zhang, Y. H. Biohydrogenation from biomass sugar mediated by *in vitro* synthetic enzymatic pathways. *Chemistry & biology* **18**, 372-380, doi:10.1016/j.chembiol.2010.12.019 (2011).

57   Krutsakorn, B. *et al.* In vitro production of n-butanol from glucose. *Metabolic engineering* **20**, 84-91, doi:10.1016/j.ymben.2013.09.006 (2013).

58   Ye, X. *et al.* Synthetic metabolic engineering-a novel, simple technology for designing a chimeric metabolic pathway. *Microbial cell factories* **11**, 120, doi:10.1186/1475-2859-11-120 (2012).

59   Morimoto, Y., Honda, K., Ye, X., Okano, K. & Ohtake, H. Directed evolution of thermotolerant malic enzyme for improved malate production. *Journal of bioscience and bioengineering*, doi:10.1016/j.jbiosc.2013.07.005 (2013).

60   Korman, T. P. *et al.* A synthetic biochemistry system for the *in vitro* production of isoprene from glycolysis intermediates. *Protein Sci* **23**, 576-585, doi:10.1002/pro.2436 (2014).

61   Bujara, M., Schumperli, M., Billerbeck, S., Heinemann, M. & Panke, S. Exploiting cell-free systems: Implementation and debugging of a system of biotransformations. *Biotechnology and bioengineering* **106**, 376-389, doi:10.1002/bit.22666 (2010).

62   Khattak, W. A. *et al.* Yeast cell-free enzyme system for bio-ethanol production at elevated temperatures. *Process Biochemistry* **49**, 357-364, doi:10.1016/j.procbio.2013.12.019 (2014).

63   Fessner, W.-D. & Walter, C. "Artificial Metabolisms" for the Asymmetric One-Pot Synthesis of Branched-Chain Saccharides. *Angewandte Chemie International Edition in English* **31**, 614-616, doi:10.1002/anie.199206141 (1992).

64   Siegal-Gaskins, D., Tuza, Z. A., Kim, J., Noireaux, V. & Murray, R. M. Gene circuit performance characterization and resource usage in a cell-free "breadboard". *ACS synthetic biology* **3**, 416-425, doi:10.1021/sb400203p (2014).

65   Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V. & Murray, R. M. Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system. *ACS synthetic biology* **3**, 387-397, doi:10.1021/sb400131a (2014).

66   Zhang, Y. H. P. Simpler Is Better: High-Yield and Potential Low-Cost Biofuels Production through Cell-Free Synthetic Pathway Biotransformation (SyPaB). *ACS Catalysis* **1**, 998-1009, doi:10.1021/cs200218f (2011).

67   You, C. *et al.* Enzymatic transformation of nonfood biomass to starch. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 7182-7187, doi:10.1073/pnas.1302420110 (2013).

68   Savile, C. K. *et al.* Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305-309, doi:10.1126/science.1188934 (2010).

69   Bechtold, M. *et al.* Biotechnological Development of a Practical Synthesis of Ethyl (S)-2-Ethoxy-3-(p-methoxyphenyl)propanoate (EEHP): Over 100-Fold Productivity Increase from Yeast Whole Cells to Recombinant Isolated Enzymes. *Organic Process Research & Development* **16**, 269-276, doi:10.1021/op200085k (2012).

70   Babich, L. *et al.* Synthesis of non-natural carbohydrates from glycerol and aldehydes in a one-pot four-enzyme cascade reaction. *Green Chemistry* **13**, 2895, doi:10.1039/c1gc15429j (2011).

71   Carlson, E. D., Gan, R., Hodgman, C. E. & Jewett, M. C. Cell-free protein synthesis: applications come of age. *Biotechnology advances* **30**, 1185-1194, doi:10.1016/j.biotechadv.2011.09.016 (2012).

72   Bujara, M., Schumperli, M., Pellaux, R., Heinemann, M. & Panke, S. Optimization of a blueprint for *in vitro* glycolysis by metabolic real-time analysis. *Nature chemical biology* **7**, 271-277, doi:10.1038/nchembio.541 (2011).

73   Harper, A. D., Bailey, C. B., Edwards, A. D., Detelich, J. F. & Keatinge-Clay, A. T. Preparative, in vitro biocatalysis of triketide lactone chiral building blocks. *Chembiochem : a European journal of chemical biology* **13**, 2200-2203, doi:10.1002/cbic.201200378 (2012).

74   Keller, M. W. *et al.* Exploiting microbial hyperthermophilicity to produce an industrial chemical, using hydrogen and carbon dioxide. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 5840-5845, doi:10.1073/pnas.1222607110 (2013).

75   Krauser, S., Weyler, C., Blass, L. K. & Heinzle, E. Directed multistep biocatalysis using tailored permeabilized cells. *Advances in biochemical engineering/biotechnology* **137**, 185-234, doi:10.1007/10_2013_240 (2013).

76   Krauser, S., Kiefer, P. & Heinzle, E. Multienzyme Whole-Cell In Situ Biocatalysis for the Production of Flaviolin in Permeabilized Cells of Escherichia coli. *ChemCatChem* **4**, 786-788, doi:10.1002/cctc.201100351 (2012).

77   Ninh, P. H. *et al.* Development of a continuous bioconversion system using a thermophilic whole-cell biocatalyst. *Applied and environmental microbiology* **79**, 1996-2001, doi:10.1128/AEM.03752-12 (2013).

78   Palmore, G. T. R., Bertschy, H., Bergens, S. H. & Whitesides, G. M. A methanol/dioxygen biofuel cell that uses NAD+-dependent dehydrogenases as catalysts: application of an electro-enzymatic method to regenerate nicotinamide adenine dinucleotide at low overpotentials. *Journal of Electroanalytical Chemistry* **443**, 155-161, doi:10.1016/s0022-0728(97)00393-8 (1998).

79   Sokic-Lazic, D. & Minteer, S. D. Pyruvate/Air Enzymatic Biofuel Cell Capable of Complete Oxidation. *Electrochemical and Solid-State Letters* **12**, F26, doi:10.1149/1.3170904 (2009).

80   Moehlenbrock, M. J., Toby, T. K., Waheed, A. & Minteer, S. D. Metabolon catalyzed pyruvate/air biofuel cell. *Journal of the American Chemical Society* **132**, 6288-6289, doi:10.1021/ja101326b (2010).

81   Minteer, S. D., Liaw, B. Y. & Cooney, M. J. Enzyme-based biofuel cells. *Current opinion in biotechnology* **18**, 228-234, doi:10.1016/j.copbio.2007.03.007 (2007).

82   Zhu, Z., Sun, F., Zhang, X. & Zhang, Y. H. Deep oxidation of glucose in enzymatic fuel cells through a synthetic enzymatic pathway containing a cascade of two thermostable dehydrogenases. *Biosensors & bioelectronics* **36**, 110-115, doi:10.1016/j.bios.2012.04.001 (2012).

83   Xu, S. & Minteer, S. D. Enzymatic Biofuel Cell for Oxidation of Glucose to CO2. *ACS Catalysis* **2**, 91-94, doi:10.1021/cs200523s (2012).

84   Myung, S. & Zhang, Y. H. Non-complexed four cascade enzyme mixture: simple purification and synergetic co-stabilization. *PloS one* **8**, e61500, doi:10.1371/journal.pone.0061500 (2013).

85   Zhu, Z., Kin Tam, T., Sun, F., You, C. & Percival Zhang, Y. H. A high-energy-density sugar biobattery based on a synthetic enzymatic pathway. *Nature communications* **5**, 3026, doi:10.1038/ncomms4026 (2014).

86   Yang, W. C., Patel, K. G., Wong, H. E. & Swartz, J. R. Simplifying and streamlining Escherichia coli-based cell-free protein synthesis. *Biotechnology progress* **28**, 413-420, doi:10.1002/btpr.1509 (2012).

87   Wang, H. H. *et al.* Multiplexed in vivo His-tagging of enzyme pathways for in vitro single-pot multienzyme catalysis. *ACS synthetic biology* **1**, 43-52, doi:10.1021/sb3000029 (2012).

88   Conrado, R. J. *et al.* DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic acids research* **40**, 1879-1889, doi:10.1093/nar/gkr888 (2012).

89   Honda, K. *et al.* Production of 2-deoxyribose 5-phosphate from fructose to demonstrate a potential of artificial bio-synthetic pathway using thermophilic enzymes. *J Biotechnol* **148**, 204-207, doi:10.1016/j.jbiotec.2010.06.008 (2010).

90   Steffler, F. & Sieber, V. Refolding of a thermostable glyceraldehyde dehydrogenase for application in synthetic cascade biomanufacturing. *PloS one* **8**, e70592, doi:10.1371/journal.pone.0070592 (2013).

91   Steffler, F., Guterl, J. K. & Sieber, V. Improvement of thermostable aldehyde dehydrogenase by directed evolution for application in Synthetic Cascade Biomanufacturing. *Enzyme and microbial technology* **53**, 307-314, doi:10.1016/j.enzmictec.2013.07.002 (2013).

92   Ye, X., Honda, K., Morimoto, Y., Okano, K. & Ohtake, H. Direct conversion of glucose to malate by synthetic metabolic engineering. *J Biotechnol* **164**, 34-40, doi:10.1016/j.jbiotec.2012.11.011 (2013).

93   Jandt, U., You, C., Zhang, Y. H. & Zeng, A. P. Compartmentalization and metabolic channeling for multienzymatic biosynthesis: practical strategies and modeling approaches. *Advances in biochemical engineering/biotechnology* **137**, 41-65, doi:10.1007/10_2013_221 (2013).

94   Obert, R. & Dave, B. C. Enzymatic Conversion of Carbon Dioxide to Methanol: Enhanced Methanol Production in Silica Sol−Gel Matrices. *Journal of the American Chemical Society* **121**, 12192-12193, doi:10.1021/ja991899r (1999).

95   Chen, A. H. & Silver, P. A. Designing biological compartmentalization. *Trends in cell biology* **22**, 662-670, doi:10.1016/j.tcb.2012.07.002 (2012).

96   Iturrate, L., Sanchez-Moreno, I., Doyaguez, E. G. & Garcia-Junceda, E. Substrate channelling in an engineered bifunctional aldolase/kinase enzyme confers catalytic advantage for C-C bond formation. *Chemical communications*, 1721-1723, doi:10.1039/b822345a (2009).

97   Wilner, O. I. *et al.* Enzyme cascades activated on topologically programmed DNA scaffolds. *Nature nanotechnology* **4**, 249-254, doi:10.1038/nnano.2009.50 (2009).

98   Lopez-Gallego, F. & Schmidt-Dannert, C. Multi-enzymatic synthesis. *Curr Opin Chem Biol* **14**, 174-183, doi:10.1016/j.cbpa.2009.11.023 (2010).

99   Dueber, J. E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nature biotechnology* **27**, 753-759, doi:10.1038/nbt.1557 (2009).

100   You, C., Myung, S. & Zhang, Y. H. Facilitated substrate channeling in a self-assembled trifunctional enzyme complex. *Angewandte Chemie* **51**, 8787-8790, doi:10.1002/anie.201202441 (2012).

101   Zhang, Y. H. Substrate channeling and enzyme complexes for biotechnological applications. *Biotechnology advances* **29**, 715-725, doi:10.1016/j.biotechadv.2011.05.020 (2011).

102   Li, Y., Su, M., Ge, X. & Tian, P. Enhanced aldehyde dehydrogenase activity by regenerating NAD+ in Klebsiella pneumoniae and implications for the glycerol dissimilation pathways. *Biotechnology letters* **35**, 1609-1615, doi:10.1007/s10529-013-1243-1 (2013).

103   Opgenorth, P. H., Korman, T. P. & Bowie, J. U. A synthetic biochemistry molecular purge valve module that maintains redox balance. *Nature communications* **5**, 4113, doi:10.1038/ncomms5113 (2014).

104   Wang, Y., San, K. Y. & Bennett, G. N. Cofactor engineering for advancing chemical biotechnology. *Current opinion in biotechnology* **24**, 994-999, doi:10.1016/j.copbio.2013.03.022 (2013).

105   Li, Y. & Cirino, P. C. Recent advances in engineering proteins for biocatalysis. *Biotechnology and bioengineering* **111**, 1273-1287, doi:10.1002/bit.25240 (2014).

106   Lee, S. H., Kwon, Y. C., Kim, D. M. & Park, C. B. Cytochrome P450-catalyzed O-dealkylation coupled with photochemical NADPH regeneration. *Biotechnology and bioengineering* **110**, 383-390, doi:10.1002/bit.24729 (2013).

107   Zheng, M. *et al.* Magnetic field intensified bi-enzyme system with in situ cofactor regeneration supported by magnetic nanoparticles. *J Biotechnol* **168**, 212-217, doi:10.1016/j.jbiotec.2013.05.016 (2013).

108   McClymont, K. & Soyer, O. S. Metabolic tinker: an online tool for guiding the design of synthetic metabolic pathways. *Nucleic acids research* **41**, e113, doi:10.1093/nar/gkt234 (2013).

109   Jensen, M. K. & Keasling, J. D. Recent applications of synthetic biology tools for yeast metabolic engineering. *FEMS Yeast Res*, doi:10.1111/1567-1364.12185 (2014).

110   Henry, C. S., Broadbelt, L. J. & Hatzimanikatis, V. Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. *Biotechnology and bioengineering* **106**, 462-473, doi:10.1002/bit.22673 (2010).

111   Gonzalez-Lergier, J., Broadbelt, L. J. & Hatzimanikatis, V. Theoretical considerations and computational analysis of the complexity in polyketide synthesis pathways. *Journal of the American Chemical Society* **127**, 9930-9938, doi:10.1021/ja051586y (2005).

112   Moriya, Y. *et al.* PathPred: an enzyme-catalyzed metabolic pathway prediction server. *Nucleic acids research* **38**, W138-143, doi:10.1093/nar/gkq318 (2010).

113   Gao, J., Ellis, L. B. & Wackett, L. P. The University of Minnesota Pathway Prediction System: multi-level prediction and visualization. *Nucleic acids research* **39**, W406-411, doi:10.1093/nar/gkr200 (2011).

114   Carbonell, P., Parutto, P., Baudier, C., Junot, C. & Faulon, J. L. Retropath: Automated Pipeline for Embedded Metabolic Circuits. *ACS synthetic biology*, doi:10.1021/sb4001273 (2013).

115   Rieckenberg, F., Ardao, I., Rujananon, R. & Zeng, A.-P. Cell-free synthesis of 1,3-propanediol from glycerol with a high yield. *Engineering in Life Sciences* **14**, 380-386, doi:10.1002/elsc.201400034 (2014).

116   Myung, S. *et al.* In vitro metabolic engineering of hydrogen production at theoretical yield from sucrose. *Metabolic engineering* **24**, 70-77, doi:10.1016/j.ymben.2014.05.006 (2014).

117   Rollin, J. A. *et al.* High-yield hydrogen production from biomass by in vitro metabolic engineering: Mixed sugars coutilization and kinetic modeling. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 4964-4969, doi:10.1073/pnas.1417719112 (2015).

118   Ninh, P. H., Honda, K., Sakai, T., Okano, K. & Ohtake, H. Assembly and multiple gene expression of thermophilic enzymes in Escherichia coli for in vitro metabolic engineering. *Biotechnology and bioengineering* **112**, 189-196, doi:10.1002/bit.25338 (2015).

119   Ullah, M. W., Khattak, W. A., Ul-Islam, M., Khan, S. & Park, J. K. Bio-ethanol production through simultaneous saccharification and fermentation using an encapsulated reconstituted cell-free enzyme system. *Biochemical Engineering Journal* **91**, 110-119, doi:10.1016/j.bej.2014.08.006 (2014).

120   Ullah, M. W., Khattak, W. A., Ul-Islam, M., Khan, S. & Park, J. K. Encapsulated yeast cell-free system: A strategy for cost-effective and sustainable production of bio-ethanol in consecutive batches. *Biotechnology and Bioprocess Engineering* **20**, 561-575, doi:10.1007/s12257-014-0855-1 (2015).

121   Kay, J. E. & Jewett, M. C. Lysate of engineered *Escherichia coli* supports high-level conversion of glucose to 2,3-butanediol. *Metabolic engineering* **32**, 133-142, doi:10.1016/j.ymben.2015.09.015 (2015).

122   Caschera, F. & Noireaux, V. A cost-effective polyphosphate-based metabolism fuels an all E. coli cell-free expression system. *Metabolic engineering* **27**, 29-37, doi:10.1016/j.ymben.2014.10.007 (2015).

123   Anderson, M. J., Stark, J. C., Hodgman, C. E. & Jewett, M. C. Energizing eukaryotic cell-free protein synthesis with glucose metabolism. *FEBS Lett* **589**, 1723-1727, doi:10.1016/j.febslet.2015.05.045 (2015).

124   Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485**, 185-194, doi:10.1038/nature11117 (2012).

125   Fritz, B. R., Timmerman, L. E., Daringer, N. M., Leonard, J. N. & Jewett, M. C. Biology by design: from top to bottom and back. *Journal of biomedicine & biotechnology* **2010**, 232016, doi:10.1155/2010/232016 (2010).

126   Rollié, S., Mangold, M. & Sundmacher, K. Designing biological systems: Systems Engineering meets Synthetic Biology. *Chemical Engineering Science* **69**, 1-29, doi:10.1016/j.ces.2011.10.068 (2012).

127   Erickson, B., Nelson & Winters, P. Perspective on opportunities in industrial biotechnology in renewable chemicals. *Biotechnology journal* **7**, 176-185, doi:10.1002/biot.201100069 (2012).

128   Demain, A. L. Importance of microbial natural products and the need to revitalize their discovery. *Journal of industrial microbiology & biotechnology* **41**, 185-201, doi:10.1007/s10295-013-1325-z (2014).

129   Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nature reviews. Drug discovery* **14**, 111-129, doi:10.1038/nrd4510 (2015).

130   Kern, A., Tilley, E., Hunter, I. S., Legisa, M. & Glieder, A. Engineering primary metabolic pathways of industrial micro-organisms. *J Biotechnol* **129**, 6-29, doi:10.1016/j.jbiotec.2006.11.021 (2007).

131   Nielsen, J. Metabolic engineering. *Appl. Microbiol. Biotechnol.* **55**, 263-283, doi:10.1007/s002530000511 (2001).

132   Green, E. M. Fermentative production of butanol--the industrial perspective. *Current opinion in biotechnology* **22**, 337-343, doi:10.1016/j.copbio.2011.02.004 (2011).

133   Lutke-Eversloh, T. & Bahl, H. Metabolic engineering of Clostridium acetobutylicum: recent advances to improve butanol production. *Current opinion in biotechnology* **22**, 634-647, doi:10.1016/j.copbio.2011.01.011 (2011).

134   Atsumi, S. *et al.* Metabolic engineering of Escherichia coli for 1-butanol production. *Metabolic engineering* **10**, 305-311, doi:10.1016/j.ymben.2007.08.003 (2008).

135   Steen, E. J. *et al.* Metabolic engineering of Saccharomyces cerevisiae for the production of n-butanol. *Microbial cell factories* **7**, 36, doi:10.1186/1475-2859-7-36 (2008).

136   Shen, C. R. *et al.* Driving forces enable high-titer anaerobic 1-butanol synthesis in Escherichia coli. *Applied and environmental microbiology* **77**, 2905-2915, doi:10.1128/AEM.03034-10 (2011).

137   Bond-Watts, B. B., Bellerose, R. J. & Chang, M. C. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nature chemical biology* **7**, 222-227, doi:10.1038/nchembio.537 (2011).

138   Dong, H. *et al.* Engineering Escherichia coli Cell Factories for n-Butanol Production. *Advances in biochemical engineering/biotechnology*, doi:10.1007/10_2015_306 (2015).

139   Dai, Z. & Nielsen, J. Advancing metabolic engineering through systems biology of industrial microorganisms. *Current opinion in biotechnology* **36**, 8-15, doi:10.1016/j.copbio.2015.08.006 (2015).

140   Lee, S. Y. & Kim, H. U. Systems strategies for developing industrial microbial strains. *Nature biotechnology* **33**, 1061-1072, doi:10.1038/nbt.3365 (2015).

141   Lee, J. W. *et al.* Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nature chemical biology* **8**, 536-546, doi:10.1038/nchembio.970 (2012).

142   Smanski, M. J. *et al.* Functional optimization of gene clusters by combinatorial design and assembly. *Nature biotechnology* **32**, 1241-1249, doi:10.1038/nbt.3063 (2014).

143   Boyle, P. M. & Silver, P. A. Parts plus pipes: synthetic biology approaches to metabolic engineering. *Metabolic engineering* **14**, 223-232, doi:10.1016/j.ymben.2011.10.003 (2012).

144   Dudley, Q. M., Karim, A. S. & Jewett, M. C. Cell-free metabolic engineering: biomanufacturing beyond the cell. *Biotechnology journal* **10**, 69-82, doi:10.1002/biot.201400330 (2015).

145   Zhang, Y. H. Production of biofuels and biochemicals by in vitro synthetic biosystems: Opportunities and challenges. *Biotechnology advances* **33**, 1467-1483, doi:10.1016/j.biotechadv.2014.10.009 (2015).

146   You, C. & Zhang, Y. H. Cell-free biosystems for biomanufacturing. *Advances in biochemical engineering/biotechnology* **131**, 89-119, doi:10.1007/10_2012_159 (2013).

147   Dodevski, I., Markou, G. C. & Sarkar, C. A. Conceptual and methodological advances in cell-free directed evolution. *Curr Opin Struct Biol* **33**, 1-7, doi:10.1016/j.sbi.2015.04.008 (2015).

148   Henrich, E., Hein, C., Dotsch, V. & Bernhard, F. Membrane protein production in Escherichia coli cell-free lysates. *FEBS Lett* **589**, 1713-1722, doi:10.1016/j.febslet.2015.04.045 (2015).

149   Zemella, A., Thoring, L., Hoffmeister, C. & Kubick, S. Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems. *Chembiochem : a European journal of chemical biology* **16**, 2420-2431, doi:10.1002/cbic.201500340 (2015).

150   Noireaux, V., Bar-Ziv, R. & Libchaber, A. Principles of cell-free genetic circuit assembly. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 12672-12677, doi:10.1073/pnas.2135496100 (2003).

151   Goshima, N. *et al.* Human protein factory for converting the transcriptome into an in vitro–expressed proteome. *Nature Methods* **5**, 1011-1017, doi:10.1038/nmeth.1273 (2008).

152   Jewett, M. C. & Swartz, J. R. Mimicking the *Escherichia coli* cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnology and bioengineering* **86**, 19-26, doi:10.1002/bit.20026 (2004).

153   Jewett, M. C. & Swartz, J. R. Substrate replenishment extends protein synthesis with an *in vitro* translation system designed to mimic the cytoplasm. *Biotechnology and bioengineering* **87**, 465-471, doi:10.1002/bit.20139 (2004).

154   Record, M. T., Courtenay, E. S., Cayley, S. & Guttman, H. J. Biophysical compensation mechanisms buffering E. coli protein–nucleic acid interactions against changing environments. *Trends in Biochemical Sciences* **23**, 190-194, doi:10.1016/s0968-0004(98)01207-9 (1998).

155   Daugherty, A. B., Govindarajan, S. & Lutz, S. Improved biocatalysts from a synthetic circular permutation library of the flavin-dependent oxidoreductase old yellow enzyme. *Journal of the American Chemical Society* **135**, 14425-14432, doi:10.1021/ja4074886 (2013).

156   Gulevich, A. Y., Skorokhodova, A. Y., Sukhozhenko, A. V., Shakulov, R. S. & Debabov, V. G. Metabolic engineering of Escherichia coli for 1-butanol biosynthesis through the inverted aerobic fatty acid beta-oxidation pathway. *Biotechnology letters* **34**, 463-469, doi:10.1007/s10529-011-0797-z (2012).

157   Nielsen, D. R. *et al.* Engineering alternative butanol production platforms in heterologous bacteria. *Metabolic engineering* **11**, 262-273, doi:10.1016/j.ymben.2009.05.003 (2009).

158   Inui, M. *et al.* Expression of Clostridium acetobutylicum butanol synthetic genes in Escherichia coli. *Appl Microbiol Biotechnol* **77**, 1305-1316, doi:10.1007/s00253-007-1257-5 (2008).

159   Yin, G. *et al.* Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system. *MAbs* **4**, 217-225, doi:10.4161/mabs.4.2.19202 (2012).

160   Zawada, J. F. *et al.* Microscale to manufacturing scale-up of cell-free cytokine production--a new approach for shortening protein production development timelines. *Biotechnology and bioengineering* **108**, 1570-1578, doi:10.1002/bit.23103 (2011).

161   Voloshin, A. M. & Swartz, J. R. Efficient and scalable method for scaling up cell free protein synthesis in batch mode. *Biotechnology and bioengineering* **91**, 516-521, doi:10.1002/bit.20528 (2005).

162   Hong, S. H. *et al.* Improving cell-free protein synthesis through genome engineering of Escherichia coli lacking release factor 1. *Chembiochem : a European journal of chemical biology* **16**, 844-853, doi:10.1002/cbic.201402708 (2015).

163   Nielsen, J. & Keasling, J. D. Engineering cellular metabolism. *Cell* **164**, 1185-1197, doi:10.1016/j.cell.2016.02.004 (2016).

164   Korman, T. P., Opgenorth, P. H. & Bowie, J. U. A synthetic biochemistry platform for cell free production of monoterpenes from glucose. *Nature communications* **8**, 15526, doi:10.1038/ncomms15526 (2017).

165   Karim, A. S., Dudley, Q. M. & Jewett, M. C. in *Industrial Biotechnology*   125-148 (2016).

166   Jensen, V. J. & Rugh, S. Industrial-scale production and application of immobilized glucose isomerase. **136**, 356-370, doi:10.1016/s0076-6879(87)36035-5 (1987).

167   Toogood, H. S. *et al.* Enzymatic menthol production: one-pot approach using engineered *Escherichia coli*. *ACS synthetic biology* **4**, 1112-1123, doi:10.1021/acssynbio.5b00092 (2015).

168   Opgenorth, P. H., Korman, T. P., Iancu, L. & Bowie, J. U. A molecular rheostat maintains ATP levels to drive a synthetic biochemistry system. *Nature chemical biology* **13**, 938-942, doi:10.1038/nchembio.2418 (2017).

169   Kwon, Y. C. & Jewett, M. C. High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci Rep* **5**, 8663, doi:10.1038/srep08663 (2015).

170   Karim, A. S., Heggestad, J. T., Crowe, S. A. & Jewett, M. C. Controlling cell-free metabolism through physiochemical perturbations. *Metabolic engineering* **45**, 86-94, doi:10.1016/j.ymben.2017.11.005 (2018).

171   Dudley, Q. M., Anderson, K. C. & Jewett, M. C. Cell-free mixing of *Escherichia coli* crude extracts to prototype and rationally engineer high-titer mevalonate synthesis. *ACS synthetic biology*, doi:10.1021/acssynbio.6b00154 (2016).

172   Hold, C., Billerbeck, S. & Panke, S. Forward design of a complex enzyme cascade reaction. *Nature communications* **7**, 12971, doi:10.1038/ncomms12971 (2016).

173   Karim, A. S. & Jewett, M. C. A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. *Metabolic engineering* **36**, 116-126, doi:10.1016/j.ymben.2016.03.002 (2016).

174   Goering, A. W. *et al. In vitro* reconstruction of nonribosomal peptide biosynthesis directly from DNA using cell-free protein synthesis. *ACS synthetic biology*, doi:10.1021/acssynbio.6b00160 (2016).

175   Du, J., Yuan, Y., Si, T., Lian, J. & Zhao, H. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic acids research* **40**, e142-e142 (2012).

176   Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **330**, 70-74 (2010).

177   Alper, H. S. & Stephanopoulos, G. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metabolic engineering* **9**, 258-267 (2007).

178   Blazeck, J., Liu, L., Redden, H. & Alper, H. S. Tuning gene expression in *Yarrowia lipolytica* by a hybrid promoter approach. *Applied and environmental microbiology* **77**, 7905-7914 (2011).

179   Kigawa, T. *et al.* Preparation of *Escherichia coli* cell extract for highly productive cell-free protein expression. *J Struct Funct Genomics* **5**, 63-68, doi:10.1023/B:JSFG.0000029204.57846.7d (2004).

180   Kim, T. W. *et al.* Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. *J Biotechnol* **126**, 554-561, doi:10.1016/j.jbiotec.2006.05.014 (2006).

181   Krinsky, N. *et al.* A simple and rapid method for preparing a cell-free bacterial lysate for protein synthesis. *PloS one* **11**, e0165137, doi:10.1371/journal.pone.0165137 (2016).

182   Shrestha, P., Holland, T. M. & Bundy, B. C. Streamlined extract preparation for *Escherichia coli*-based cell-free protein synthesis by sonication or bead vortex mixing. *Biotechniques* **53**, 163-174, doi:10.2144/0000113924 (2012).

183   Sun, Z. Z. *et al.* Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *J Vis Exp*, e50762, doi:10.3791/50762 (2013).

184   Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nature biotechnology* **23**, 102-107, doi:10.1038/nbt1044 (2005).

185   Spedding, G. *Ribosomes and protein synthesis, a practical approach*. 318 (Oxford University Press, 1990).

186   Erlacher, M. D., Chirkova, A., Voegele, P. & Polacek, N. Generation of chemically engineered ribosomes for atomic mutagenesis studies on protein biosynthesis. *Nat Protoc* **6**, 580-592, doi:10.1038/nprot.2011.306 (2011).

187   Polacek, N. The ribosome meets synthetic biology. *Chembiochem : a European journal of chemical biology* **12**, 2122-2124, doi:10.1002/cbic.201100259 (2011).

188   Forster, A. C. & Church, G. M. Towards synthesis of a minimal cell. *Mol Syst Biol* **2**, 45, doi:10.1038/msb4100090 (2006).

189   Jewett, M. C. & Forster, A. C. Update on designing and building minimal cells. *Current opinion in biotechnology* **21**, 697-703, doi:10.1016/j.copbio.2010.06.008 (2010).

190   Cochella, L. & Green, R. Isolation of antibiotic resistance mutations in the rRNA by using an in vitro selection system. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 3786-3791, doi:10.1073/pnas.0307596101 (2004).

191   Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441-444, doi:10.1038/nature08817 (2010).

192   Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature biotechnology* **25**, 770-777, doi:10.1038/nbt1314 (2007).

193   Traub, P. N., M. Structure and function of E. coli ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. *Proceedings of the National Academy of Sciences of the United States of America* **59**, 8 (1968).

194   Maki, J. A. & Culver, G. M. Recent developments in factor-facilitated ribosome assembly. *Methods* **36**, 313-320, doi:10.1016/j.ymeth.2005.04.008 (2005).

195   Nierhaus, K. D., F. Total reconstitution of functionally active 50S ribosomal subunits from Escherichia coli. *Proc. Nat. Acad. Sci. USA* **71**, 5 (1974).

196   Green, R. & Noller, H. F. In vitro complementation analysis localizes 23S rRNA posttranscriptional modifications that are required for Escherichia coli 50S ribosomal subunit assembly and function. . *RNA* **2**, 1011--1021 (1996).

197   Semrad, K. & Green, R. Osmolytes stimulate the reconstitution of functional 50S ribosomes from in vitro transcripts of Escherichia coli 23S rRNA. *RNA* **8**, 401--411 (2002).

198   Fritz, B. R., Jamil, O. K. & Jewett, M. C. Implications of macromolecular crowding and reducing conditions for in vitro ribosome construction. *Nucleic acids research* **43**, 4774-4784, doi:10.1093/nar/gkv329 (2015).

199   Fritz, B. R. & Jewett, M. C. The impact of transcriptional tuning on in vitro integrated rRNA transcription and ribosome construction. *Nucleic acids research* **42**, 6774-6785, doi:10.1093/nar/gku307 (2014).

200   Jewett, M. C., Fritz, B. R., Timmerman, L. E. & Church, G. M. In vitro integration of ribosomal RNA synthesis, ribosome assembly, and translation. *Mol Syst Biol* **9**, 678, doi:10.1038/msb.2013.31 (2013).

201   Liu, Y., Fritz, B. R., Anderson, M. J., Schoborg, J. A. & Jewett, M. C. Characterizing and alleviating substrate limitations for improved in vitro ribosome construction. *ACS synthetic biology* **4**, 454-462, doi:10.1021/sb5002467 (2015).

202   Caschera, F. & Noireaux, V. Integration of biological parts toward the synthesis of a minimal cell. *Curr Opin Chem Biol* **22**, 85-91, doi:10.1016/j.cbpa.2014.09.028 (2014).

203   Adamala, K. P., Martin-Alarcon, D. A., Guthrie-Honea, K. R. & Boyden, E. S. Engineering genetic circuit interactions within and between synthetic minimal cells. *Nat Chem* **9**, 431-439 (2016).

204   Caschera, F., Lee, J. W., Ho, K. K., Liu, A. P. & Jewett, M. C. Cell-free compartmentalized protein synthesis inside double emulsion templated liposomes with in vitro synthesized and assembled ribosomes. *Chemical communications* **52**, 5467-5469, doi:10.1039/c6cc00223d (2016).

205   Calhoun, K. A. & Swartz, J. R. Energy systems for ATP regeneration in cell-free protein synthesis reactions. *Methods Mol Biol* **375**, 3-17, doi:10.1007/978-1-59745-388-2_1 (2007).

206   Caschera, F. Bacterial cell-free expression technology to in vitro systems engineering and optimization. *Synth Syst Biotechnol* **2**, 97-104, doi:10.1016/j.synbio.2017.07.004 (2017).

207   Caschera, F. & Noireaux, V. Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system. *Biochimie* **99**, 162-168, doi:10.1016/j.biochi.2013.11.025 (2014).

208   Caschera, F. & Noireaux, V. A cost-effective polyphosphate-based metabolism fuels an all E. coli cell-free expression system. *Metabolic Engineering* **27**, 29-37, doi:10.1016/j.ymben.2014.10.007 (2015).

209   Kurylo, C. M. *et al.* Genome Sequence and Analysis of Escherichia coli MRE600, a Colicinogenic, Nonmotile Strain that Lacks RNase I and the Type I Methyltransferase, EcoKI. *Genome Biol Evol* **8**, 742-752, doi:10.1093/gbe/evw008 (2016).

210   Forlin M. *et al.* Evolutionary experiments for self-assembling amphiphilic systems. *Chemometrics and Intelligent Laboratory Systems* **90**, 153-160 (2008).

211   Caschera, F. *et al.* Automated discovery of novel drug formulations using predictive iterated high throughput experimentation. *PloS one* **5**, e8546, doi:10.1371/journal.pone.0008546 (2010).

212   Caschera, F. *et al.* Coping with complexity: machine learning optimization of cell-free protein synthesis. *Biotechnology and bioengineering* **108**, 2218-2228, doi:10.1002/bit.23178 (2011).

213   Wang, Y. & Zhang, Y. H. Cell-free protein synthesis energized by slowly-metabolized maltodextrin. *BMC Biotechnol* **9**, 58, doi:10.1186/1472-6750-9-58 (2009).

214   Calhoun, K. A. & Swartz, J. R. An economical method for cell-free protein synthesis using glucose and nucleoside monophosphates. *Biotechnology progress* **21**, 1146-1153, doi:10.1021/bp050052y (2005).

215   Soye, B. J. D., Patel, J. R., Isaacs, F. J. & Jewett, M. C. Repurposing the translation apparatus for synthetic biology. *Curr Opin Chem Biol* **28**, 83-90, doi:10.1016/j.cbpa.2015.06.008 (2015).

216   Wen, K. Y. *et al.* A Cell-Free Biosensor for Detecting Quorum Sensing Molecules in P. aeruginosa-Infected Respiratory Samples. *ACS synthetic biology* **6**, 2293-2301, doi:10.1021/acssynbio.7b00219 (2017).

217   Garamella, J., Marshall, R., Rustad, M. & Noireaux, V. The All E-coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS synthetic biology* **5**, 344-355, doi:10.1021/acssynbio.5b00296 (2016).

218   Kim, T. W. *et al.* High-throughput in vitro glycoside hydrolase (HIGH) screening for enzyme discovery. *Angewandte Chemie* **50**, 11215-11218, doi:10.1002/anie.201104685 (2011).

219   Mak, W. S. *et al.* Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway. *Nature communications* **6**, 10005, doi:10.1038/ncomms10005 (2015).

220   Kightlinger, W. *et al.* Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases. *Nat Chem Biol*, doi:10.1038/s41589-018-0051-2 (2018).

221   Hong, S. H. *et al.* Improving Cell-Free Protein Synthesis through Genome Engineering of Escherichia coli Lacking Release Factor 1. **16**, 844-853 (2015).

222   Martin, R. W. *et al.* Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. *Nat Commun* **9**, 1203, doi:10.1038/s41467-018-03469-5 (2018).

223   Oza, J. P. *et al.* Robust production of recombinant phosphoproteins using cell-free protein synthesis. *Nature communications* **6**, 8168, doi:ARTN 8168

10.1038/ncomms9168 (2015).

224   Caschera, F. & Noireaux, V. Preparation of amino acid mixtures for cell-free expression systems. *Biotechniques* **58**, 40-43, doi:10.2144/000114249 (2015).

225   Shin, J. & Noireaux, V. Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. *Journal of biological engineering* **4**, 8, doi:10.1186/1754-1611-4-8 (2010).

226   Lietzke, R. & Nierhaus, K. H. Total reconstitution of 70S ribosomes from Escherichia coli. *Methods Enzymol* **164**, 278-283 (1988).

227   Culver, G. M. & Noller, H. F. Efficient reconstitution of functional Escherichia coli 30S ribosomal subunits from a complete set of recombinant small subunit ribosomal proteins. *RNA* **5**, 832-843 (1999).

228   Matsubayashi, H. & Ueda, T. Purified cell-free systems as standard parts for synthetic biology. *Curr Opin Chem Biol* **22**, 158-162, doi:10.1016/j.cbpa.2014.09.031 (2014).

229   Fothergillgilmore, L. A. The Evolution of the Glycolytic Pathway. *Trends in Biochemical Sciences* **11**, 47-&, doi:Doi 10.1016/0968-0004(86)90233-1 (1986).

230   Spirin, A. S. High-throughput cell-free systems for synthesis of functionally active proteins. *Trends in biotechnology* **22**, 538-545, doi:10.1016/j.tibtech.2004.08.012 (2004).

231   Alissandratos, A., Caron, K., Loan, T. D., Hennessy, J. E. & Easton, C. J. ATP Recycling with Cell Lysate for Enzyme-Catalyzed Chemical Synthesis, Protein Expression and PCR. *ACS Chem Biol* **11**, 3289-3293, doi:10.1021/acschembio.6b00838 (2016).

232   Kim, T. W. *et al.* Prolonged cell-free protein synthesis using dual energy sources: Combined use of creatine phosphate and glucose for the efficient supply of ATP and retarded accumulation of phosphate. *Biotechnology and bioengineering* **97**, 1510-1515, doi:10.1002/bit.21337 (2007).

233   Failmezger, J., Nitschel, R., Sanchez-Kopper, A., Kraml, M. & Siemann-Herzberg, M. Site-Specific Cleavage of Ribosomal RNA in Escherichia coli-Based Cell-Free Protein Synthesis Systems. *PloS one* **11**, e0168764, doi:10.1371/journal.pone.0168764 (2016).

234   Caschera, F. & Noireaux, V. Compartmentalization of an all-E. coli Cell-Free Expression System for the Construction of a Minimal Cell. *Artif Life* **22**, 185-195, doi:10.1162/ARTL_a_00198 (2016).

235   Kuem, J. W., Kim, T. W., Park, C. G., Choi, C. Y. & Kim, D. M. Oxalate enhances protein synthesis in cell-free synthesis system utilizing 3-phosphoglycerate as energy source. *Journal of bioscience and bioengineering* **101**, 162-165, doi:10.1263/jbb.101.162 (2006).

236   Moore, S. J. *et al.* Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria. *Proc Natl Acad Sci U S A* **115**, E4340-E4349, doi:10.1073/pnas.1715806115 (2018).

237   Marshall, R. *et al.* Rapid and Scalable Characterization of CRISPR Technologies Using an E. coli Cell-Free Transcription-Translation System. *Mol Cell* **69**, 146-157 e143, doi:10.1016/j.molcel.2017.12.007 (2018).

238   Luisi, P. L., Ferri, F. & Stano, P. Approaches to semi-synthetic minimal cells: a review. *Naturwissenschaften* **93**, 1-13, doi:10.1007/s00114-005-0056-z (2006).

239   Chizzolini, F., Forlin, M., Cecchi, D. & Mansy, S. S. Gene Position More Strongly Influences Cell-Free Protein Expression from Operons than T7 Transcriptional Promoter Strength. *ACS synthetic biology* **3**, 363-371, doi:10.1021/sb4000977 (2014).

240   Paddon, C. J. & Keasling, J. D. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat Rev Microbiol* **12**, 355-367, doi:10.1038/nrmicro3240 (2014).

241   Chang, A. *et al.* BRENDA in 2015: exciting developments in its 25th year of existence. *Nucleic acids research* **43**, D439-446, doi:10.1093/nar/gku1068 (2015).

242   Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic acids research* **42**, D459-471, doi:10.1093/nar/gkt1103 (2014).

243   Na, D., Kim, T. Y. & Lee, S. Y. Construction and optimization of synthetic pathways in metabolic engineering. *Curr Opin Microbiol* **13**, 363-370, doi:10.1016/j.mib.2010.02.004 (2010).

244   Morgado, G., Gerngross, D., Roberts, T. M. & Panke, S. Synthetic Biology for Cell-Free Biosynthesis: Fundamentals of Designing Novel In Vitro Multi-Enzyme Reaction Networks. *Advances in biochemical engineering/biotechnology*, doi:10.1007/10_2016_13 (2016).

245   Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature chemical biology* **5**, 593-599, doi:10.1038/nchembio.186 (2009).

246   Garamella, J., Marshall, R., Rustad, M. & Noireaux, V. The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS synthetic biology* **5**, 344-355, doi:10.1021/acssynbio.5b00296 (2016).

247   Takahashi, M. K. *et al.* Characterizing and prototyping genetic networks with cell-free transcription–translation reactions. *Methods* **86**, 60-72, doi:https://doi.org/10.1016/j.ymeth.2015.05.020 (2015).

248   Kelwick, R., Webb, A. J., MacDonald, J. T. & Freemont, P. S. Development of a Bacillus subtilis cell-free transcription-translation system for prototyping regulatory elements. *Metabolic engineering* **38**, 370-381, doi:10.1016/j.ymben.2016.09.008 (2016).

249   Moore, S. J., Lai, H. E., Needham, H., Polizzi, K. M. & Freemont, P. S. Streptomyces venezuelae TX-TL - a next generation cell-free synthetic biology tool. *Biotechnology journal* **12**, doi:10.1002/biot.201600678 (2017).

250   Calhoun, K. A. & Swartz, J. R. Energizing cell-free protein synthesis with glucose metabolism. *Biotechnology and bioengineering* **90**, 606-613, doi:10.1002/bit.20449 (2005).

251   Caschera, F. & Noireaux, V. Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription–translation system. *Biochimie* **99**, 162-168, doi:https://doi.org/10.1016/j.biochi.2013.11.025 (2014).

252   Knapp, K. G., Goerke, A. R. & Swartz, J. R. Cell-free synthesis of proteins that require disulfide bonds using glucose as an energy source. *Biotechnology and bioengineering* **97**, 901-908, doi:10.1002/bit.21296 (2007).

253   Yu, L., Xu, M., Tang, I. C. & Yang, S. T. Metabolic engineering of Clostridium tyrobutyricum for n-butanol production through co-utilization of glucose and xylose. *Biotechnology and bioengineering* **112**, 2134-2141, doi:10.1002/bit.25613 (2015).

254   Dusséaux, S., Croux, C., Soucaille, P. & Meynial-Salles, I. Metabolic engineering of Clostridium acetobutylicum ATCC 824 for the high-yield production of a biofuel composed of an isopropanol/butanol/ethanol mixture. *Metabolic engineering* **18**, 1-8, doi:https://doi.org/10.1016/j.ymben.2013.03.003 (2013).

255   Davidi, D. *et al.* Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro kcat measurements. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 3401-3406, doi:10.1073/pnas.1514240113 (2016).

256   Wu, Y. Y. *et al.* System-level studies of a cell-free transcription-translation platform for metabolic engineering. *BioRxiv* **Preprint**, doi:10.1101/172007 (2017).

257   Kelwick, R. *et al.* Cell-free prototyping strategies for enhancing the sustainable production of polyhydroxyalkanoates bioplastics. *BioRxiv* **Preprint**, doi:10.1101/225144 (2017).

258   Karim, A. S. & Jewett, M. C. Cell-Free Synthetic Biology for Pathway Prototyping. *Methods in Enzymology*
, doi:10.1016/bs.mie.2018.04.029 (2018).

259   Bumpus, S. B., Evans, B. S., Thomas, P. M., Ntai, I. & Kelleher, N. L. A proteomics approach to discovering natural products and their biosynthetic pathways. *Nature biotechnology* **27**, 951-956, doi:10.1038/nbt.1565 (2009).

260   Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Natural product reports* **18**, 380-416 (2001).

261   Koehn, F. E. & Carter, G. T. The evolving role of natural products in drug discovery. *Nature reviews. Drug discovery* **4**, 206-220, doi:10.1038/nrd1657 (2005).

262   Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412-421, doi:10.1016/j.cell.2014.06.034 (2014).

263   Curtis, T. P., Sloan, W. T. & Scannell, J. W. Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10494-10499, doi:10.1073/pnas.142680199 (2002).

264   Dykhuizen, D. Species Numbers in Bacteria. *Proceedings. California Academy of Sciences* **56**, 62-71 (2005).

265   Nguta, J. M., Appiah-Opong, R., Nyarko, A. K., Yeboah-Manu, D. & Addo, P. G. A. Current perspectives in drug discovery against tuberculosis from natural products. *International Journal of Mycobacteriology* **4**, 165-183, doi:10.1016/j.ijmyco.2015.05.004 (2015).

266   Medema, M. H. & Fischbach, M. A. Computational approaches to natural product discovery. *Nature chemical biology* **11**, 639-648, doi:10.1038/nchembio.1884 (2015).

267   Weber, T. *et al.* antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic acids research* **43**, W237-243, doi:10.1093/nar/gkv437 (2015).

268   Mayer, E. A., Savidge, T. & Shulman, R. J. Brain-gut microbiome interactions and functional bowel disorders. *Gastroenterology* **146**, 1500-1512, doi:10.1053/j.gastro.2014.02.037 (2014).

269   Clarke, G. *et al.* The microbiome-gut-brain axis during early life regulates the hippocampal serotonergic system in a sex-dependent manner. *Mol Psychiatry* **18**, 666-673, doi:10.1038/mp.2012.77 (2013).

270   Moloney, R. D., Desbonnet, L., Clarke, G., Dinan, T. G. & Cryan, J. F. The microbiome: stress, health and disease. *Mamm Genome* **25**, 49-74, doi:10.1007/s00335-013-9488-5 (2014).

271   Donia, M. S. *et al.* A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**, 1402-1414, doi:10.1016/j.cell.2014.08.032 (2014).

272   Robey, M. T. *et al.* Identification of the First Diketomorpholine Biosynthetic Pathway Using FAC-MS Technology. *ACS Chem Biol*, doi:10.1021/acschembio.8b00024 (2018).

273   Guo, C. J. *et al.* Discovery of Reactive Microbiota-Derived Metabolites that Inhibit Host Proteases. *Cell* **168**, 517-526 e518, doi:10.1016/j.cell.2016.12.021 (2017).

Appendix A: Supplementary Information

Table 0.1. Strains and Plasmids. Strains used for plasmid cloning, protein expression, and extract preparation are listed in the top portion of the table. Plasmids used for CFPS are listed in the bottom portion.

Name	Genotype/relevant characteristics	Source
Strains		
NEB Turbo™	F' proA+B+ lacIq ΔlacZM15 / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)TetS endA1 thi-1 Δ(hsdS-mcrB)5	New England Biolabs
BL21 (DE3)	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5	New England Biolabs
Strains & Plasmids		
BL21 (DE3) pETBCS-rbsU-atoB	Strain used for protein production and extract preparation of AtoB	This Study
BL21 (DE3) pETBCS-rbsU-hbd1	Strain used for protein production and extract preparation of Hbd1	This Study
BL21 (DE3) pETBCS-rbsU-hbd2	Strain used for protein production and extract preparation of Hbd2	This Study
BL21 (DE3) pETBCS-rbsU-crt1	Strain used for protein production and extract preparation of Crt1	This Study
BL21 (DE3) pETBCS-rbsU-crt2	Strain used for protein production and extract preparation of Crt2	This Study
BL21 (DE3) pETBCS-rbsU-ter1	Strain used for protein production and extract preparation of Ter1	This Study
BL21 (DE3) pETBCS-rbsU-adhE1	Strain used for protein production and extract preparation of AdhE1	This Study
BL21 (DE3) pETBCS-rbsU-adhE2	Strain used for protein production and extract preparation of AdhE2	This Study
pJL1-atoB	Plasmid used for CFPS containing atoB	This Study
pJL1-hbd2	Plasmid used for CFPS containing hbd2	This Study
pJL1-crt1	Plasmid used for CFPS containing crt1	This Study
pJL1-ter1	Plasmid used for CFPS containing ter1	This Study
pJL1-adhE1	Plasmid used for CFPS containing adhE1	This Study
pJL1-hbdcrt2	Plasmid used for CFPS containing hbdcrt2	This Study
pJL1-hbdcrt3	Plasmid used for CFPS containing hbdcrt3	This Study
pJL1-hbdcrt4	Plasmid used for CFPS containing hbdcrt4	This Study
pJL1-hbdcrt6	Plasmid used for CFPS containing hbdcrt6	This Study
pJL1-ter3	Plasmid used for CFPS containing ter3	This Study

pJL1-ter4	Plasmid used for CFPS containing ter4	This Study
pJL1-ter5	Plasmid used for CFPS containing ter5	This Study
pJL1-ter6	Plasmid used for CFPS containing ter6	This Study
pJL1-adhE9	Plasmid used for CFPS containing adhE9	This Study
pJL1-adhE10	Plasmid used for CFPS containing adhE10	This Study
pJL1-adhE13	Plasmid used for CFPS containing adhE13	This Study

Table 0.2. Genes and Enzymes. *Each gene, corresponding enzyme, and source organism used in this study is listed.*

Gene/ Enzyme	Enzyme Activity	Source Organism	Source Database
atoB	acetyl-CoA acetyltransferase / thiolase	*Escherichia coli*	GenBank
hbd1	acyl-CoA dehydrogenase	*Clostridium Acetobutylicum*	GenBank
hbd2	acyl-CoA dehydrogenase	*Clostridium beijerinckii*	GenBank
crt1	acyl-CoA dehydrogenase	*Clostridium Acetobutylicum*	GenBank
crt2	acyl-CoA dehydrogenase	*Pseudomonas putida*	GenBank
hbdcrt2	3-hydroxyacyl-CoA dehydrogenase	*Aeropyrum camini*	NCBI BLAST Search of ADHE2 protein from C. acetobutyllicum
hbdcrt3	3-hydroxyacyl-CoA dehydrogenase	*Pyrobaculum aerophilum*	NCBI BLAST Search of ADHE2 protein from C. acetobutyllicum
hbdcrt4	3-hydroxyacyl-CoA dehydrogenase	*Sulfolobus islandicus*	NCBI BLAST Search of ADHE2 protein from *C. acetobutyllicum*
hbdcrt5	3-hydroxyacyl-CoA dehydrogenase	*Vulcanisaeta distributa*	NCBI BLAST Search of TER protein from T. denticola
hbdcrt6	3-hydroxybutyryl-CoA dehydrogenase	*Sulfolobus acidocaldarius*	NCBI BLAST Search of TER protein from T. denticola
ter1	trans-2-enoyl-CoA reductase	*Treponema denticola*	GenBank
ter3	trans-2-enoyl-CoA reductase	*Fibrobacter succinogenes*	NCBI BLAST Search of TER protein from T. denticola
ter4	trans-2-enoyl-CoA reductase	*Flavobacterium johnsoniae*	NCBI BLAST Search of TER protein from T. denticola
ter5	trans-2-enoyl-CoA reductase	*Spirochaeta bajacaliforniensis*	NCBI BLAST Search of TER protein from T. denticola
ter6	trans-2-enoyl-CoA reductase	*Cytophaga hutchinsonii*	NCBI BLAST Search of HBD-CRT protein from M. sedula
adhE1	bifunctional acetaldehyde-CoA/alcohol dehydrogenase	*Clostridium acetobutylicum*	GenBank
adhE2	bifunctional acetaldehyde-CoA/alcohol dehydrogenase	Clostridium pasteurianum	NCBI BLAST Search of HBD-CRT protein from M. sedula
adhE8	bifunctional acetaldehyde-CoA/alcohol dehydrogenase	*Chitinivibrio alkaliphilus*	NCBI BLAST Search of HBD-CRT protein from M. sedula
adhE9	bifunctional acetaldehyde-CoA/alcohol dehydrogenase	*Thermosynechococcus sp. NK55a*	NCBI BLAST Search of HBD-CRT protein from M. sedula
adhE10	bifunctional acetaldehyde-CoA/alcohol dehydrogenase	*Providencia burhodogranariea*	NCBI BLAST Search of HBD-CRT protein from M. sedula

| *adhE13* | bifunctional acetaldehyde-CoA/alcohol dehydrogenase | *Serratia marcescens* | NCBI BLAST Search of HBD-CRT protein from M. sedula |

Table 0.3. Primers and Plasmid Construct Descriptions. *Primers used to make constructs in* **Table 0.2** *are listed with descriptions of their use.*

Name	Description	Primers	
pETBCS-rbsU	PCR primers for the plasmid backbone	Fwd	CTCGAGCACCACCACCAC
		Rev	TGCGGATCCTCCTTATGTG
pETBCS-rbsU-atoB	PCR primers to add 20 basepair homology to the pET backbone for Gibson Assembly of atoB into pET	Fwd	ATAAGGAGGATCCGCAATGAAAAATTGTG TCATCG
		Rev	TCAGTGGTGGTGGTGGTGGTGCTCGAGTT AATTCAACCGTTCAATCAC
pETBCS-rbsU-hbd1	PCR primers to add 20 basepair homology to the pET backbone for Gibson Assembly of hbd1 into pET	Fwd	ATAAGGAGGATCCGCAATGAAAAAGGTGT GCGTG
		Rev	TCAGTGGTGGTGGTGGTGGTGCTCGAGTT ATTTGCTATAATCGTAGAAACC
pETBCS-rbsU-hbd2	PCR primers to add 20 basepair homology to the pET backbone for Gibson Assembly of hbd2 into pET	Fwd	ATAAGGAGGATCCGCAATGAAAAAGATTT TTGTGTTG
		Rev	TCAGTGGTGGTGGTGGTGGTGCTCGAGTT ATTTGCTGTAATCATAGAAGC
pETBCS-rbsU-crt1	PCR primers to add 20 basepair homology to the pET backbone for Gibson Assembly of crt1 into pET	Fwd	ATAAGGAGGATCCGCAATGGAACTTAATA ACGTAATCTTAGAAAAG
		Rev	TCAGTGGTGGTGGTGGTGGTGCTCGAGT CAGCGGTTTTTGAATCC
pETBCS-rbsU-crt2	PCR primers to add 20 basepair homology to the pET backbone for Gibson Assembly of crt2 into pET	Fwd	ATAAGGAGGATCCGCAATGACAACCCCGA GCAGC
		Rev	TCAGTGGTGGTGGTGGTGGTGCTCGAGT CACTGGCCGGTGAACCG
pETBCS-rbsU-ter1	PCR primers to add 20 basepair homology to the pET backbone for Gibson Assembly of ter1 into pET	Fwd	ATAAGGAGGATCCGCAATGATCGTGAAAC CTATGGTG
		Rev	TCAGTGGTGGTGGTGGTGGTGCTCGAGTT AGATCCGGTCGAACCG
pETBCS-rbsU-adhE1	PCR primers to add 20 basepair homology to the pET backbone for Gibson Assembly of adhE1 into pET	Fwd	ACACATAAGGAGGATCCGCAATGAAAGTG ACCAACCAGAAAG
		Rev	CAGTGGTGGTGGTGGTGGTGCTCGAGTT AAAAGCTTTTAATATAGATATCCTTTAATTC G
pETBCS-rbsU-adhE2	PCR primers to add 20 basepair homology to the pET backbone for Gibson Assembly of adhE2 into pET	Fwd	ATAAGGAGGATCCGCAATGAAAGTGTCGA ACGTG
		Rev	TCAGTGGTGGTGGTGGTGGTGCTCGAGTT ATTTAATTTCGGTGTCAATTTC
pJL1	PCR primers for the plasmid backbone	Fwd	GTCGACCGGCTGCTAACA
		Rev	ATGTATATCTCCTTCTTAAAGTTAAACAAA ATTATTTC
pJL1-atoB	PCR primers to add 20 basepair homology to the pJL1 backbone for Gibson Assembly of atoB into pJL1	Fwd	TTTAAGAAGGAGATATACATATGAAAAATT GTGTCATCG
		Rev	TTTGTTAGCAGCCGGTCGACTTAATTCAAC CGTTCAATCAC
pJL1-hbd1	PCR primers to add 20 basepair homology to the pJL1 backbone for Gibson Assembly of hbd1 into pJL1	Fwd	TTTAAGAAGGAGATATACATATGAAAAAGG TGTGCGTG
		Rev	TTTGTTAGCAGCCGGTCGACTTATTTGCTA TAATCGTAGAAACC
pJL1-hbd2	PCR primers to add 20 basepair homology to the pJL1	Fwd	TTTAAGAAGGAGATATACATATGAAAAAGA TTTTTGTGTTG

	backbone for Gibson Assembly of hbd2 into pJL1	Rev	TTTGTTAGCAGCCGGTCGACTTATTTGCT GTAATCATAGAAGC
pJL1-crt1	PCR primers to add 20 basepair homology to the pJL1	Fwd	TTTAAGAAGGAGATATACATATGGAACTTA ATAACGTAATCTTAGAAAAG
	backbone for Gibson Assembly of crt1 into pJL1	Rev	TTTGTTAGCAGCCGGTCGACTCAGCGGTT TTTGAATCC
pJL1-crt2	PCR primers to add 20 basepair homology to the pJL1	Fwd	TTTAAGAAGGAGATATACATATGACAACCC CGAGCAGC
	backbone for Gibson Assembly of crt2 into pJL1	Rev	TTTGTTAGCAGCCGGTCGACTCACTGGCC GGTGAACCG
pJL1-hbdcrt1	PCR primers to add 20 basepair homology to the pJL1	Fwd	TTTAAGAAGGAGATATACATATGAAAGTCA CGGTCATCGG
	backbone for Gibson Assembly of hbdcrt1 into pJL1	Rev	TTTGTTAGCAGCCGGTCGACTCACTCGCC GCGGAAGTT
pJL1-hbdcrt2	PCR primers to add 20 basepair homology to the pJL1	Fwd	TTTAAGAAGGAGATATACATATGGCTGGC GAAGTGAAA
	backbone for Gibson Assembly of hbdcrt2 into pJL1	Rev	TTTGTTAGCAGCCGGTCGACTTAGCGGCC CTTAAAACG
pJL1-hbdcrt3	PCR primers to add 20 basepair homology to the pJL1	Fwd	TTTAAGAAGGAGATATACATATGCCAAAAG TGGCGGTG
	backbone for Gibson Assembly of hbdcrt3 into pJL1	Rev	TTTGTTAGCAGCCGGTCGACTTATTTGCC CTTGAACTGCG
pJL1-hbdcrt4	PCR primers to add 20 basepair homology to the pJL1	Fwd	TTTAAGAAGGAGATATACATATGAAAGTTG AGGATATTAAAAAG
	backbone for Gibson Assembly of hbdcrt4 into pJL1	Rev	TTTGTTAGCAGCCGGTCGACTTATTCGCC TTTAAACTGC
pJL1-hbdcrt5	PCR primers to add 20 basepair homology to the pJL1	Fwd	TTTAAGAAGGAGATATACATATGTCGTCGT TGGGAATTAAG
	backbone for Gibson Assembly of hbdcrt5 into pJL1	Rev	TTTGTTAGCAGCCGGTCGACTCATTCTCC GGTAAAACG
pJL1-hbdcrt6	PCR primers to add 20 basepair homology to the pJL1	Fwd	TTTAAGAAGGAGATATACATATGGAGATCC GTAAATTCAC
	backbone for Gibson Assembly of hbdcrt6 into pJL1	Rev	TTTGTTAGCAGCCGGTCGACTCATTTATAC TCGTAGAATCCTTTAC
pJL1-ter1	PCR primers to add 20 basepair homology to the pJL1	Fwd	TTTAAGAAGGAGATATACATATGATCGTGA AACCTATGGTG
	backbone for Gibson Assembly of ter1 into pJL1	Rev	TTTGTTAGCAGCCGGTCGACTTAGATCCG GTCGAACCG
pJL1-ter3	PCR primers to add 20 basepair homology to the pJL1	Fwd	TTTAAGAAGGAGATATACATATGATTATTA AACCACTGATCC
	backbone for Gibson Assembly of ter3 into pJL1	Rev	TTTGTTAGCAGCCGGTCGACTCAAATTGA GGTTAACGTTTG
pJL1-ter4	PCR primers to add 20 basepair homology to the pJL1	Fwd	TTTAAGAAGGAGATATACATATGATCATCG AGCCGCGC
	backbone for Gibson Assembly of ter4 into pJL1	Rev	TTTGTTAGCAGCCGGTCGACTTATTTGATG CTTTCAATGTTTACGACC
pJL1-ter5	PCR primers to add 20 basepair homology to the pJL1	Fwd	TTTAAGAAGGAGATATACATATGGAACCG AAGATTCGC
	backbone for Gibson Assembly of ter5 into pJL1	Rev	TTTGTTAGCAGCCGGTCGACTCAAATGTC GAGGAAGTCC
pJL1-ter6	PCR primers to add 20 basepair homology to the pJL1	Fwd	TTTAAGAAGGAGATATACATATGATTATCG AGCCCAAAATG

	backbone for Gibson Assembly of ter6 into pJL1	Rev	TTTGTTAGCAGCCGGTCGACTTAAACCAGGCCTTCAATG
pJL1-adhE1	PCR primers to add 20 basepair homology to the pJL1 backbone for Gibson Assembly of adhE1 into pJL1	Fwd	TTTAAGAAGGAGATATACATATGAAAGTGACCAACCAG
		Rev	TTTGTTAGCAGCCGGTCGACTCCTAAAAGCTTTTAATATAGATATCC
pJL1-adhE2	PCR primers to add 20 basepair homology to the pJL1 backbone for Gibson Assembly of adhE2 into pJL1	Fwd	TTTAAGAAGGAGATATACATATGAAAGTGTCGAACGTG
		Rev	TTTGTTAGCAGCCGGTCGACTTATTTAATTTCGGTGTCAATTTC
pJL1-adhE9	PCR primers to add 20 basepair homology to the pJL1 backbone for Gibson Assembly of adhE9 into pJL1	Fwd	TTTAAGAAGGAGATATACATATGAACGCTCCGACGCTG
		Rev	TTTGTTAGCAGCCGGTCGACTTAGCTGGCCGGAGCTTC
pJL1-adhE10	PCR primers to add 20 basepair homology to the pJL1 backbone for Gibson Assembly of adhE10 into pJL1	Fwd	TTTAAGAAGGAGATATACATATGAGCGTCACTAATGTGAC
		Rev	TTTGTTAGCAGCCGGTCGACTTATTTTTTGGTTGATTTTTTTTC
pJL1-adhE13	PCR primers to add 20 basepair homology to the pJL1 backbone for Gibson Assembly of adhE13 into pJL1	Fwd	TTTAAGAAGGAGATATACATATGGCGGTCACCAACGTG
		Rev	TTTGTTAGCAGCCGGTCGACTTATTTCTTAGGTTTCTTAACGGCCG
All Linear Templates	PCR primers to linearize the pJL1 regulatory elements and inserted gene with an added randomized GC-rich ~20bp for use in CFPS	Fwd	ATGCAGGTCATCCGAGGGGTTCGGATCCCGCGAAATTAATACG
		Rev	GCTCGCGGATCACTCGTCACTCCGCTTTCAGCAAAAAACCCCTC

Table 0.4. Codon Optimized Gene Sequences for Chapter 3. *The sequences of genes that were codon optimized by IDT's optimizer are listed.*

Gene	Sequence
hbd1	ATGAAAAAGGTGTGCGTGATTGGTGCGGGTACCATGGGTAGCGGCATCGCGCAAGCG TTTGCGGCGAAAGGCTTTGAAGTGGTGCTGCGCGATATCAAAGATGAATTTGTGGATC GCGGCCTGGATTTTATCAACAAAAACTTAAGCAAACTGGTGAAAAAAGGCAAGATTGAA GAAGCGACCAAAGTGGAAATCCTGACCCGCATCTCAGGCACCGTGGACTTGAACATG GCGGCGGATTGTGATCTGGTGATTGAAGCGGCGGTGGAACGCATGGATATCAAAAAG CAAATCTTTGCGGACCTGGACAACATTTGTAAGCCGGAAACCATCTTAGCGAGCAACA CCAGCAGCTTGAGCATTACCGAAGTAGCGAGCGCGACCAAAACCAACGATAAGGTGA TTGGTATGCATTTCTTTAACCCGGCGCCGGTGATGAAGTTAGTGGAGGTGATTCGCGG CATTGCGACCAGCCAGGAAACCTTTGATGCGGTGAAAGAGACGAGCATTGCGATTGG CAAAGATCCGGTGGAAGTGGCGGAAGCGCCGGGCTTTGTGGTGAACCGCATTCTGAT TCCGATGATCAACGAAGCGGTGGGTATTCTGGCGGAAGGCATTGCGAGCGTGGAAGA CATTGATAAAGCGATGAAATTAGGCGCGAACCACCCGATGGGCCCGCTGGAACTGGG TGATTTTATTGGTTTAGATATTGCTTGGCGATTATGGATGTGCTGTACAGCGAAACCG GCGATAGCAAGTATCGCCCGCATACCCTGTTGAAGAAGTATGTGCGCGCGGGCTGGT TGGGCCGCAAAAGCGGCAAAGGTTTCTACGATTATAGCAAATAA
hbd2	ATGAAAAAGATTTTTGTGTTGGGCGCGGGCACCATGGGTGCGGGTATCGTGCAGGCG TTCGCGCAGAAAGGTTGCGAAGTGATCGTGCGCGACATTAAGGAAGAATTTGTGGACC GCGGCATTGCGGGCATCACCAAAGGCCTGGAAAAGCAGGTGGCGAAAGGCAAAATGA GCGAAGAAGATAAAGAAGCGATTTTTAAGCCGCATCAGCGGCACCACCGATATGAAACT GGCGGCGGACTGCGATCTGGTGGTGGAAGCGGCGATCGAAAATATGAAAATCAAGAA GGAAATCTTCGCGGAACTGGATGGCATCTGCAAGCCGGAAGCTATCCTGGCGAGCAA TACCAGCAGCCTGAGCATCACCGAAGTGGCGAGCGCGACCAAGCGCCCGGATAAAGT GATCGGCATGCATTTCTTTAACCCGGCGCCGGTGATGAAGTTGGTGGAAATCATCAAA GGCATTGCGACCAGCCAGGAAACCTTTGATGCGGTGAAGGAACTGAGCGTGGCGATC GGCAAAGAACCGGTGGAAGTGGCGGAAGCGCCGGGCTTCGTGGTGAATCGCATTCTG ATCCCGATGATCAATGAAGCGAGCTTTATCTTACAGGAAGGCATTGCGAGCGTGGAAG ATATCGATACCGCGATGAAATATGGTGCGAATCATCCGATGGGCCCGCTGGCGCTGG GCGATTTGATCGGCCTGGACGTGTGTCTGGCGATCATGGATGTGCTGTTCACCGAAAC CGGTGATAATAAGTACCGCGCGTCATCAATTCTGCGCAAATATGTGCGCGCGGGCTG GTTGGGCCGCAAAAGCGGCAAAGGCTTCTATGATTACAGCAAATAA
crt1	ATGGAACTTAATAACGTAATCTTAGAAAAGGAAGGTAAAGTGGCGGTGGTGACAATCA ATCGCCCGAAAGCGCTGAACGCTCTGAACAGCGATACCCTCAAAGAAATGGATTATGT GATTGGTGAAATCGAAACGATTCAGAAGTGTTAGCGGTGATCCTGACCGGCGCGGG CGAAAAAAGCTTTGTGGCGGGCGCGGATATCAGCGAGATGAAGGAAATGAACACAAT CGAAGGTCGCAAATTCGGAATTTTGGGCAACAAAGTATTTCGCCGCCTGGAATTATTA GAAAGCCGGTGATTGCGGCGGTGAACGGTTTTGCGCTGGGCGGAGGCTGTGAAATT GCGATGAGCTGCGATATTCGCATTGCGAGCTCAAATGCGCGCTTTGGTCAGCCGGAA GTGGGTCTAGGCATTACCCCGGGTTTTGGTGGTACCCAGCGCTTAAGCCGCCTGGTG GGCATGGGAATGGCGAAGCAATTGATTTTTACCGCGCAGAACATTAAGGCGGATGAAG CGCTGCGCATCGGCTTGGTGAACAAGGTGGTGGAACCGAGCGAACTGATGAACACCG CGAAAGAAATCGCGAATAAAATCGTATCAAACGCGCCGGTGGCGGTGAAGCTGTCAAA ACAAGCGATCAACCGCGGCATGCAATGCGATATCGATACCGCGCTGGCGTTTGAAAG CGAAGCGTTTGGCGAATGTTTTAGCACCGAGGATCAGAAGGATGCGATGACCGCGTT CATTGAGAAACGCAAAATCGAAGGATTCAAAAACCGCTGA
crt2	ATGACAACCCCGAGCAGCCCTCTGTTAAGCAAAGTTGAGGCTGGCGTAGCGTGGATTA CCTTGAACCGCCCAGAACAGCGCAACGCCCTGGATATCCCAACCTTAAAACAACTGCA TGCGTTATTAGATAGCCACGCGGATGATCCAGCGGTACGCGTGGTGGTGCTGACCGG CAGCGGCCGCAGCTTTTGCGCTGGCGCGGATCTGGCGGAGTGGGCTGCGGCGGAGG CTGCGGGCACCCTGGAGAGCTACGGCTGGACCGAGACAGCGCACGCGCTGATGTTG CGCCTGCATAGCTTGGATAAGCCAACCATTGCGGCGATTAACGGCACCGCGGTGGGC GGGGGCATGGATCTCAGCCTGTGCTGCGATCTGCGCATTGCGGCGGCGAGCGCCCG

```
         CTTTAAAGCGGGCTATACCAGCATGGGCTATAGCCCAGACGCGGGCGCGAGCTGGCA
         TCTGCCTCGGCTGATTGGCAGCGAACAGGCGAAACGCTTGTTATTTTTGGACGAGCTG
         TGGGGCGCGGAACACGCGCTGGCCGCTGGGCTGGTTAGCGAGGTTTGCGCGGATGA
         ACAACTGCCAGCGGTGGCCGCGGAATTAGCGGGGCGCCTGGCGAATGGCCCGACTT
         TTGCGTACGCCCAGACCAAACAGCTGATTCGCGATGGCGCGCGGCGCACCTTAGCGG
         AACAGCTGGAAGCTGAACGCCATGCGGGCCTGCTGTGCGGCCGCAGCCAGGACGGC
         GCGGAAGCGCTGCAAGCGAGCGTAGAGCGCCGCGCGCCACGGTTCACCGGCCAGTG
         A
ter1     ATGATCGTGAAACCTATGGTGCGGAATAACATCTGCCTCAATGCCCATCCACAGGGTT
         GTAAAAAAGGTGTGGAGGATCAGATCGAGTACACGAAAAAGCGTATCACGGCCGAGG
         TGAAAGCCGGTGCGAAAGCCCCGAAAAATGTACTTGTGCTCGGTTGTTCCAATGGTTA
         CGGTCTGGCGTCTCGGATCACTGCGGCGTTCGGTTATGGAGCCGCCACTATCGGTGT
         GTCGTTTGAAAAAGCCGGTAGTGAAACGAAGTACGGTACGCCAGGTTGGTACAATAAT
         CTCGCCTTTGACGAGGCCGCCAAACGGGAAGGTCTGTATTCCGTGACTATCGACGGT
         GACGCCTTTTCCGATGAAATCAAGGCGCAGGTTATCGAAGAAGCCAAAAAGAAAGGTA
         TCAAATTCGATCTGATCGTGTACTCCTTAGCGTCGCCGGTGCGTACGGACCCAGACAC
         GGGTATCATGCACAAGTCCGTATTGAAACCATTCGGTAAAACGTTTACGGGTAAGACG
         GTAGATCCATTTACTGGTGAACTCAAAGAAATCTCTGCCGAACCGGCGAATGATGAAG
         AAGCCGCCGCGACGGTGAAGGTGATGGGTGGTGAAGACTGGGAACGGTGGATCAAG
         CAACTGTCCAAAGAAGGTCTGTTGGAAGAAGGTTGTATCACGCTCGCCTATTCTTATAT
         CGGTCCGGAGGCCACTCAGGCCCTCTACCGGAAGGGTACGATCGGTAAGGCGAAAG
         AACACCTGGAAGCGACTGCCCACCGGCTCAATAAAGAAAATCCATCCATCCGGGCCTT
         CGTGAGTGTAAATAAGGGTTTGGTGACGCGGGCGTCAGCTGTGATCCCAGTGATCCC
         ACTCTACCTCGCCAGTTTGTTTAAGGTGATGAAAGAAAAAGGTAATCACGAGGGTTGC
         ATCGAGCAAATCACGCGGCTGTATGCCGAACGTCTTTACCGGAAAGACGGTACGATCC
         CAGTAGATGAAGAAATCGTATCCGGATCGACGACTGGGAGCTGGAAGAAGACGTGC
         AGAAGGCTGTGTCGGCCTTGATGGAAAAAGTGACGGGTGAAAACGCCGAATCCCTCA
         CGGACCTGGCGGGTTACCGGCACGACTTTTTGGCCTCCAATGGTTTTGATGTGGAAG
         GTATCAATTACGAAGCCGAGGTGGAACGGTTCGACCGGATCTAA
adhE1    ATGAAGGTAACCAACCAGAAAGAGCTGAAACAAAAATTAAACGAACTCCGCGAAGCGC
         AAAAAAAATTCGCGACGTATACTCAGGAACAAGTCGATAAGATCTTTAAACAATGTGCC
         ATTGCAGCGGCCAAAGAACGCATCAACCTGGCGAAGTTGGCCGTTGAAGAAACCGGA
         ATTGGTTTAGTGGAAGACAAAATTATTAAGAACCATTTCGCTGCGGAATATATTTATAAT
         AAATACAAAAATGAGAAGACCTGCGGAATTATTGATCATGATGATAGCCTTGGTATCAC
         TAAAGTAGCAGAACCAATCGGTATCGTCGCCGCCATCGTTCCTACAACCAATCCGACC
         TCTACGGCGATCTTTAAATCATTGATTAGCCTGAAAACGCGTAACGCGATTTTTTTCAG
         CCCTCACCCACGCGCCAAAAAAAGCACTATCGCTGCGGCGAAACTGATTCTGGATGC
         GGCAGTTAAAGCCGGCGCACCTAAAAACATTATCGGCTGGATCGACGAGCCTAGCATC
         GAGTTGAGCCAGGACCTCATGAGTGAAGCAGATATTATCCTCGCCACGGGTGGGCCA
         TCTATGGTTAAAGCGGCCTACTCATCTGGTAAACCAGCCATCGGTGTGGGTGCGGGCA
         ATACCCCGGCGATCATTGACGAGAGCGCCGATATTGATATGGCCGTTAGTAGCATCAT
         TCTGAGCAAAACCTACGATAACGGCGTAATTTGCGCGAGTGAACAGAGCATTTTAGTG
         ATGAACTCGATCTATGAAAAAGTGAAAGAAGAATTTGTGAAGCGCGGTTCTTACATCCT
         CAACCAAAATGAAATCGCGAAAATCAAAGAAACGATGTTCAAAAATGGCGCGATCAAC
         GCGGATATTGTTGGCAAATCAGCCTACATTATTGCGAAAATGGCGGGTATTGAAGTCC
         CCCAGACCACAAAGATCCTGATCGGTGAAGTACAGAGCGTCGAAAAGAGCGAGCTGT
         TCAGCCACGAGAAACTGAGCCCTGTTCTGGCCATGTACAAGGTAAAAGATTTTGACGA
         AGCACTTAAAAAAGCCCAACGCCTTATCGAATTAGGAGGGTCTGGCCACACGAGCAGC
         TTGTACATCGACAGCCAGAATAACAAAGACAAAGTCAAAGAATTCGGCCTTGCAATGAA
         AACTTCTCGCACCTTTATTAATATGCCGTCCAGCCAGGGCGCCTCTGGTGATCTGTAC
         AATTTTGCCATTGCCCCGTCGTTTACCCTGGGGGTGTGGGACCTGGGGCGGGAATTCG
         GTATCACAGAACGTCGAACCAAAACACCTGTTGAATATTAAATCCGTGGCAGAGCGCC
         GCGAGAACATGCTGTGGTTCAAAGTCCCTCAGAAAATTTACTTCAAGTACGGCTGCCT
         GCGTTTTGCGCTGAAAGAACTCAAAGACATGAACAAAAACGTGCGTTCATCGTTACC
         GATAAGGACCTGTTTAAACTGGGCTACGTAAACAAAATTACTAAAGTGTTGGACGAAAT
```

CGATATTAAATACTCCATTTTTACCGACATTAAGTCAGACCCGACCATCGACAGCGTCA
AAAAAGGGGCAAAAGAAATGCTGAACTTTGAGCCAGATACGATTATCTCAATCGGTGG
GGGCTCGCCTATGGACGCTGCGAAAGTGATGCACCTGCTGTATGAGTACCCGGAAGC
GGAGATTGAGAACCTGGCCATCAATTTTATGGATATTCGTAAGCGTATTTGCAATTTTC
CGAAACTTGGGACGAAAGCCATCTCCGTCGCGATTCCGACCACTGCAGGTACGGGCA
GCGAAGCCACGCCTTTTGCAGTTATCACGAACGATGAGACCGGTATGAAATATCCGCT
CACCTCGTACGAACTGACCCCAAATATGGCCATCATTGATACCGAACTGATGCTCAATA
TGCCCCGTAAACTCACCGCAGCCACTGGCATTGACGCACTCGTGCACGCCATTGAGG
CTTATGTCAGCGTGATGGCGACCGATTACACCGATGAATTAGCTCTCCGTGCAATCAA
AATGATTTTTAAGTACCTCCCGCGTGCGTACAAAAATGGCACGAATGATATCGAGGCG
CGTGAAAAGATGGCTCATGCCAGCAACATCGCCGGTATGGCGTTCGCTAATGCCTTCC
TGGGTGTATGCCACAGTATGGCACACAAGCTCGGCGCCATGCATCATGTACCACACG
GGATTGCCTGTGCCGTGTTAATCGAGGAAGTCATTAAGTACAATGCCACCGATTGCCC
GACTAAACAGACCGCCTTTCCGCAGTATAAGAGCCCGAATGCAAAACGTAAATACGCC
GAGATCGCTGAGTATTTGAATCTCAAAGGAACGAGTGATACTGAGAAAGTCACCGCCT
TGATCGAAGCCATCAGCAAACTTAAAATCGATCTGTCGATTCCGCAGAACATTAGCGC
GGCCGGTATTAACAAGAAAGATTTTTACAACACGCTGGACAAAATGTCTGAACTGGCG
TTTGACGATCAGTGCACCACGGCGAACCCGCGCTATCCTCTGATTTCCGAGCTCAAGG
ATATCTACATCAAAAGCTTCTAA

adhE2   ATGAAAGTGTCGAACGTGGACGAATTGAACTTGCGTCTGCAGGAGATGCGTACGGCC
CAGGAAAGTTTGCAGCGTACACTCAGGAACAGGTCGACGAAATTTTCCGTCAGGCTG
CTATGGCCGCGCTGGATGCACGCATCCCCCTGGCCAAGATGACTGTCGAAGAGACAG
GGATGGGCCTGGTTGAGGACAAAGTGATCAAAAACCATTTCGCCGCGGAATACATTTA
TAACCAATACAAAGACGAAAAAACTTGCGGTGTGATCGAAACGGATGAAAGCTATGGA
ATTACGAAAATCGCGGAGCCAATTGGGATTGTGGCCGCGGTGATTCCGACGACGAAT
CCGACATCCACGGCGATTTTCAAAACGCTGATTGCCCTGAAGACCCGTAACGCGATCA
TGCTCTCTCCTCATCCCCGTGCTAAAAAACTCAACCATCGCCGCGGCAAAAATTATCCTG
GATGCCGCCGTTAAAGCTGGTGCACCCGAAGGCATTATCGGCTGGATCGATGAACCG
TCCATCGAACTGACACAAATCCTGATGCAGGAAGCTGACATCACCCTCGCAACCGGCG
GCCCCAGTATGGTGAAGAGCGCGTATTCGTCTGGGAAACCAGCCATCGGCGTGGGCC
CGGGTAACACCCCGATTATCATCGATGAGAGTGCCCATATCAAGATGGCAGTGTCTTC
TATCATTCTTAGCAAAACCTTTGATAACGGCGTAATTTGTGCTAGCGAACAGTCTGTTAT
CGTACTGGATAGCATCTACGACAAAGTCCGCAAAGAATTTGCCGAACGCGGGGCATAT
ATCATCAAACCAGATGAGATGGATAAGGTCCGCAATACTATTTTCATTAATGGGTCGAT
TAACGCAAAAATTGTGGGTCAAAGCGCGCCCAAATCGCGGAAATGGCCGGCATTAAG
GTGCTTGAAACGGCACGTATCCTGATCGGGGAAGTCACCTCATTGACCGAGGAAGAA
GCATTCGCGCATGAAAAACTGTCTACCGTCCTGGCGATGTATCGCGCCGAAAATTTCG
AGGATGCGCTGGAAAAAGCGGTGACGCTGGTTAATTTAGGCGGTCTTGGTCATACCAG
CGGGATCTATGCGGACGTCGTGAAAGCGAAAGACCGCATTGACAAATTTAGTTCGGCT
ATGAAAACAGTCCGTACGTTCATCAACATCCCAACCGCCCAAGGCGCAAGCGGAGATC
TCTATAACTTTAAAATCGCACCTAGTTTCACGCTGGGATGTGGTTCATGGGGAGGGAA
CAGCGTGAGCGAAAACGTTGGGCCCAAGCACCTTCTCAATATCAAACGCGTCGCGGA
GCGTCGCGAAAATATGCTGTGGTTCCGTGTGCCTGAGAAAGTGTACTTCAAATTTGGC
TGCCTGCAATTTGCGCTCCGCGAGCTGAAAGATTTAAACAAGAAACGTGCGTTCATTG
TAACGGACAAAGTGCTGTATGACCTGGGATACGCGGACACGATCACGAAAGTTCTGGA
GGAGATCGGAGTTGATTTCAAAGTCTTTACGGAGGTGGAACCTGACCCAACTCTTGCT
ACCGCACGTAAAGGTACAGAGGAAATGATGGATTTTAAACCAGATACTATCATTTCGCT
GGGAGGGGGGTTCTGCAATGGACGCGGCGAAGATCATGTGGGTGTTATACGAACATCC
GGAGGTGAAATTTGAGGATTTGGCGATGCGCTTTATGGACATTCGCAAACGTATTTACA
ACTTCCCGAAACTGGGGGGAGAAAGCCATGATGATCGCTGTCGCTACTTCGGCGGGCA
CAGGTTCCGAAGTGACACCTTTTGCGGTGATCACCGATGAAAAGACTGGTGTGAAATA
TCCGCTGGCCGATTATGAATTAACCCCAGATATGGCAATCGTAGATGCTGAATTGATGA
TGAATATGCCAAAGGGCCTGACAGCTGCTAGCGGTATTGACGCTTTGATCCATGGTAT
CGAAGCCTACACGTCGGTATTAGCCTCAGAGTATACTAACGGCCTCGCGCTCGAAGCC
ATTCGCCTGATTTTTAAATATCTCCCTGAGGCATACTCGGAAGGTATCACCAATGAGAA

AGCACGTGAAAAAATGGCGCATGCATCCACTATCGCCGGTATGGCCTTTGCAAACGCC
TTCTTGGGTGTGTGCCATTCAATGGCGCATAAACTCGGAGCGGAGCACCATATTGCCC
ATGGCACTGCTAACGCTCTGCTGATCGAAGAAGTCATCCGCTTCAACGCAGTTGACAA
CCCAGTGAAACAGGCGGCATTCCCGCAATATAAATATCCGAATGCGAAATGGCGCTAT
GGAAAAATTGCGGATTACCTGAAGCTTGGCGGCTCAACCGATGAAGAAAAAGTGGAAT
TGCTCATCAAAGCCATCCATGAGTTGAAAGAAAAGGTCAATTTACCAATGACCATTAAA
GAGGCGGGTGTTAGCGAGAAAAAATTTTACGCTACCCTCGACAAGATGTGTGAACTCG
CTTTTGACGATCAGTGCACCGGTGCGAACCCGCGCTACCCGCTGATCAGCGAGATCA
AACAGATGTTTATCAACGCATTCGACAAGGCCGAAATTGACACCGAAATTAAATAA

ter3   ATGATTATTAAACCACTGATCCGCTCTAATATGTGTATCAACGCGCATCCGAAAGGTTG
TGCCGCCGACGTGAAACATCAAATCGAGTTCATCAAAAAGAAATTCACGACCCGCTCA
ATCCCGGCGGACGCGCCAAAAACAGTGTTAGTCCTGGGCTGCTCCACTGGATACGGC
TTAGCATCACGCATCGTCGCGGCTTTTGGTTACAAGGCTGCAACGATTGGGGTATCGT
TCGAAAAGAAGGCTCCGACGGAGGAATCGGTGAGAGTCGTGAGAAAACAGGCACCC
CGGGCTGGTATAACAACATGGCGTTTGATAAGTTCGCGAAGGAAGCCGGTCTGGATG
CGGTCACCTTCAACGGTGACGCCTTTAGCCATGAAATGCGTCAGAATGTTATCGATAC
CCTGAAAAAAATGGGTCGCAAAGTAGATCTCTTGGTCTATTCTGTCGCAAGCTCAGTCC
GCGTTGATCCAGATAACGGGACCATCTACCGCTCAGTTCTGAAGCCCATCGACAAAGT
GTTCACCGGGGCGACGATCGATTGCCTGTCTGGTAAGATTTCGACAATTTCGGCCGAA
CCTGCGACGGCAGAAGAAGCGGCGAACACGGTCAAAGTGATGGGTGGCGAGGATTG
GGCGTTGTGGGTGCGCAAACTGAAAGAGGCAGGCGTCCTTGCGGAAGGTGTTAAAAC
TGTGGCCTATTCCTATATCGGCCCGAAACTCAGCCACGCTATCTATCGCGACGGCACT
ATCGGGGGTGCCAAAAAACACTTGGAAGCTACGGCTCTTGAACTTAACAAAGAGCTCC
AGAATGATCTCCATGGGGAGGCGTATGTGTCGGTGAATAAAGGTTTAGTGACGCGCAG
CTCAGCAGTGATCCCGATCATTCCGATGTACATTTCGGTTCTGTTTAAAGTCATGAAAG
AAATGGGCAACCACGAAGGCTGTATTGAACAGATGGAACGCCTGATGACGGAACGCTT
GTATACCGGCTCTAAAGTGCCCACCGACGAAAACCATTTGATCCGTATTGACGATTATG
AATTGGATCCGAAGGTCCAGGCGGAAGTTGATAAGCGCATGGCTACAGTGACTCAGG
AAAATTTGGCGGAAGTGGGTGATCTGGAAGGATACCGTCACGACTTTTTGGCAACCAA
TGGCTTCGATATTGACGGTGTGGACTACGAGGCCGATGTGCAAACGTTAACCTCAATT
TGA

ter4   ATGATCATCGAGCCGCGCATGCGCGGTTTTATCTGCCTGACTGCGCATCCGGCGGGA
TGTGAACAGAATGTTAAAAATCAGATCGAGTATATTAAATCGAAAGGGGCAATCGCCG
GCGCCAAAAAGGTTCTGGTGATCGGCGCATCCACGGGTTTCGGTTTAGCATCCCGTAT
CACCAGTGCGTTCGGCTCAGATGCTGCTACGATTGGCGTGTTCTTCGAAAAACCGCCC
GTCGAAGGTAAGACAGCGTCGCCAGGGTGGTATAATTCGGCCGCATTTGAGAAAGAG
GCACATAAAGCGGGTCTTTACGCTAAATCTATCAATGGAGACGCTTTCAGCAACGAAAT
TAAACGTGAAACCTTAGATCTGATCAAAGCGGATTTAGGTCAGGTTGATCTGGTAATTT
ATTCGCTGGCGTCCCCGGTTCGTACGAACCCGAACACAGGTGTGACTCACCGCAGTG
TGTTGAAACCGATCGGTCAGACTTTTACAAACAAAACTGTGGATTTTCATACGGGGAAC
GTGTCCGAAGTTTCTATCGCGCCGGCTAATGAAGAAGATATTGAAAATACGGTAGCAG
TGATGGGCGGAGAAGATTGGGCGATGTGGATTGATGCCCTCAAAAATGAAATCTGCT
GGCAGAGGGGGCGACGACAATTGCATATTCCTATATTGGCCCGGAATTGACCGAAGC
GGTCTACCGTAAAGGCACCATTGGTCGTGCAAAAGACCACCTGGAGGCGACCGCTTT
CACCATTACTGATACCCTTAAATCGTTAGGCGGAAAAGCGTACGTGTCGGTGAATAAA
GCCTTGGTTACGCAAGCCTCGTCGGCGATTCCTGTGATCCCGCTGTATATCTCGCTGC
TTTATAAAATTATGAAGGAGGAAGGAATTCACGAGGGATGCATCGAACAAATTCAGCG
CTTGTTCCAAGATCGTTTGTATAACGGTAGCGAAGTGCCGGTTGATGAGAAAGGCCGC
ATCCGCATTGACGATTGGGAGATGCGCGAGGATGTGCAGGCTAAAGTTGCGGCTCTG
TGGAAGGAAGCCACCACCGAAACCCTGCCATCCATCGGCGACCTGGCAGGTTACCGT
AATGACTTCTTAAACCTGTTTGGGTTTGAATTTGCGGGAGTGGATTACAAGGCGGATAC
GAACGAGGTCGTAAACATTGAAAGCATCAAATAA

ter5   ATGGAACCGAAGATTCGCGGCGCCATTTGTATGAACGCCCACCCAAAAGGCTGTGCA
GAAGAAACGCGTCGCCAGATCCACTGGACGCAGGACTATTTCAGCCATCATAAAGCGT
CTAAGAAACTCAAGCGCGTTCTGGTAATTGGTGCGAGTACTGGTTACGGCCTTGCGAG

TCGTATTGCCGCTGCCTTTGGCTATGGTGCCGCCACAATCGGCGTCAGCTTCGAAAAA
GAACCGAGCGAGAAGCGCCCTGCCACCCCGGGTTGGTACAACACTAAGGCTTTTGAT
GAAGAAGCAAAGAAAGCGGGGGATCCCGGCAGTAAGCTTTAATGGAGACGCCTTCAGC
AACGAAATGCGTGCTAAGGTTGGTGACGCGTTGAAATCCCTCGGTGGGCCAGCGGAT
TTAGTCATTTACAGCCTGGCGAGCGGCGTGCGCACCGATCCGGCGGACGGGACCCTG
TATCGTAGCGTTCTTAAACCTCTCGGTGAAGTTTATAAAGCAAAGAGCGTTGACTTCAT
TAACGGGCGCCTCGGTGAAGTGGAAATTGCTCCAGCGACAGAAGAAGAGCGTCGTGC
CACGGTGAAAGTTATGGGTGGTGAAGATTGGCGTCTGTGGATTGACTATTTGGGCGCG
CAAGGGTTATTAGCTCCGGATGTTAAACGCTTGCGTATAGTTATATCGGCCCGGAAG
TTACCCACGCCGTGTACCGTGAAGGCACCATCGGGAGCGCAAAGAAAGATCTGGAAG
AGACGGCTAAACGCCTGGACAAGGAACTGAGTGCCAGCGGGGGTGGCGCGTACGTA
AGCGTGAATAAAGCTGTTGTGACTCGTGCATCGGCGGTGATCCCGGTTGTTTCGCTGT
ACATTAGTCTTCTGTTCAAAGTTATGAAGCAAAAAGGGTTACACGAGGGATGTATTGAA
CAGATGGTACGTCTTTTTTTCTCAGCGTCTGTACTCGGATAAAGCAGTTAGCGTCGACG
CAGAAGGGCGTATTCGCATCGATGATTGGGAAATGCGTGAAGACATCCAGGAAGAAGT
AAGCTCGCTGTGGGAGCGCGTAGACGACGAAAATATCGAAGAACTCGCGGACCTGGC
AGGGTATCGCGAGAACTTCCTTAAAATCCACGGCTTTTCTGTACCAGGTATTGATTATG
ACGCAGAAGTGGACTTCCTCGACATTTGA

ter6   ATGATTATCGAGCCCAAAATGCGCGGCTTCATTTGCCTCACCTCTCACCCTACAGGCT
GCGAGCAAAACGTGATTAACCAGATTAACTATGTCAAATCGAAAGGCGTTATCAACGG
CCCTAAAAAAGTGCTGGTGATCGGTGCGAGTACCGGTTTTGGCCTTGCGAGTCGTATT
ACTTCGGCTTTTGGTTCCAACGCGGCCACCATTGGGGTGTTTTTCGAAAAACCAGCAC
AGGAAGGGAAGCCGGGAAGCCCGGGCTGGTATAACACCGTAGCGTTCCAGAACGAA
GCCAAGAAGGCTGGTATCTATGCAAAAGCATCAACGGTGATGCCTTTTCCACTGAGG
TGAAGCAGAAAACCATCGACCTGATTAAGGCAGATTTAGGTCAAGTGGATCTTGTTATT
TACTCCCTCGCAAGCCCGGTGCGTACCAACCCCGTCACCGGCGTTACACACCGTAGC
GTCTTAAAACCTATCGGTGGTGCCTTCTCAAACAAAACCGTAGATTTTCATACGGGTAA
TGTTAGCACTGTGACGATCGAACCGGCGAATGAGGAGGATGTTACCAACACTGTAGCG
GTGATGGGGGGCGAAGATTGGGGCATGTGGATGGATGCGATGTTGGAAGCTGGAGTC
CTGGCTGAAGGTGCGACCACCGTGGCTTATTCTTACATTGGCCCAGCCCTGACAGAAG
CCGTATACCGCAAAGGTACGATTGGCCGTGCAAAAGACCATCTGGAAGCGAGTGCAG
CCACCATTACCGATAAACTTAAAAGTGTGAAAGGAAAAGCCTACGTGTCTGTGAACAAG
GCGCTGGTGACCCAGGCGTCGTCGGCAATTCCTGTAATCCCGCTGTATATCAGTTTGC
TGTATAAGGTTATGAAGGCCGAAGGGATTCATGAAGGCTGCATCGAGCAAATCCAACG
CTTGTATGCAGACCGCTTATATACTGGTAAAGCGATCCCGACTGACGAACAGGGCCGT
ATCCGTATTGATGATTGGGAAATGCGCGAAGATGTCCAGGCGAACGTCGCCGCACTGT
GGGAACAGGTTACGAGTGAAAACGTTAGCGATATCTCCGACTTGAAAGGTTACAAAAA
CGATTTCCTGAACCTGTTTGGCTTTGCAGTGAACAAAGTGGACTATCTGGCTGACGTTA
ACGAGAACGTGACCATTGAAGGCCTGGTTTAA

adhE9   ATGAACGCTCCGACGCTGAACTCATACCCCCCTGTTCAGAGCCTTGCAGATCTTGAGG
GCCTGATTGAACGCGTGCAACGCGCCCAGTCCCAGTATGCGCAGTTCACGCAAGAAC
AGGTTGATCACATCTTCCATCAGGCAGCAATGGCGGCTAATCAAGCACGTATTCCTCT
CGCTAAGCAGGCCGTAGCGGAGACCGGTATGGGCGTCGTGGAAGATAAAGTAATTAA
GAACCATTTTGCTAGCGAATATATTTACAATAAATACAAAAATGAAAAAACGTGTGGTGT
GATTGAGGATGATCCCATTTTTGGCATCCAAAAAATTGCGGAACCGGTGGGTATTATC
GCAGGGGTTGTGCCGGTTACGAACCCTACTTCAACAACCATCTTCAAAGCGTTAATTG
CTCTTAAGACGCGCAATGGGATTATTTTTTCTCCGCACCCTCGTGCAAAAGGGTGCAC
CGTCGCCGCGGCAAAAGTGGTGCTGGATGCCGCCGTTGCGGCGGGTGCGCCGCCGG
ATATTATCGGATGGATTGATGAACCCACGATCGAATTATCCCAGGCCCTGATGCAGCA
TCCACAGATCAAACTGATCCTGGCCACTGGCGGTAGCGGTATGGTTAAAGCCGCCTAT
TCTAGCGGGCATCCTGCCATCGGTGTGGGCGCAGGCAACACCCCTGTGTTGATCGAT
GCGACTGCCGATATCCCGACCGCGGTTTCTTCTATCTTGCTTTCTAAGGCGTTCGATAA
TGGTATGATTTGCGCCTCGGAACAGGCTGTTATTGTCGTGGATGAAATTTATGATGCC
GTAAAAGCCGAATTCCAGCGTCGCGGTGCCTATATTCTGTCACCTGAAGAACGTCAGC
AGGTGGCGCAGTTGCTGCTGAAAGACGGGCGCCTGAATGCGGCCATCGTTGGTCAAT

CGGCCGCAACCATTGCAGCCATGGCGAACATCCAAGTGCCGCCGCAAACGCGCGTTC
TGATTGGTGAAGTGTCTGAAGTGGGCCCGCAGGAGCCGTTCTCGTACGAAAAGCTGT
GCCCGGTGCTGGCTTTATATCGCGCGCCACAGTTTCATAAAGGAGTGGAAATTGCCGC
GCAATTAGTTAATTTCGGTGGTAAAGGCCATACGTCGGTTCTGTACACCGACCCGCGT
AACCAGGATGACATCGCCTATTTTAAATATCGCTTACAAACGGCGCGCGTCTTAATTAA
TACCCCGAGTTCACAGGGTGCCATTGGCGACCTGTATAACTTTAAACTGGATCCTAGC
CTGACACTGGGCTGTGGCACTTGGGGGGGTAACGTCACTTCAGAAAACGTAGGCCCG
CGTCATCTCTTGAATATCAAAACCGTAAGTGATCGTCGTGAAAATATGTTGTGGTTTCG
CGTGCCGCCAAAAATCTACTTCAAACCGGGCTGTCTGCCAATTGCACTGCGCGAATTG
GCAGGCAAAAGCGCGCATTTCTGGTGACCGACCAGCCGTTATTCGACCTGGGAATC
ACCGAACCGATCGTGCATACACTGGCAGAGCTCGACATTAAATACGATATTTTTCACGA
AGTTGAACCCGACCCGACGCTGAGCACAGTTAAACGCGGGCTGGAACTGTTACGCCA
GTACCAGCCGGATGTAATTATCGCTGTTGGAGGTGGTTCCCCCATGGATGCAGCCAAA
GTTATGTGGCTTCTCTACGAACATCCGACTGTTGAATTCGACGGTCTTGCGATGCGCTT
TATGGATATTCGTAAACGTGTCTATCAATTGCCTCCGCTGGGCCAAAAAGCGATCCTC
GTTGCCATTCCGACGACCAGTGGTACGGGGTCGGAAGTGACCCCGTTTGCCGTAGTT
ACCGATGATCGCGTGGGTATTAAATATCCTCTCGCGGACTATGCACTGACCCCAACTA
TGGCGATTGTTGACCCTGATCTGGTACTCCACATGCCGAAAAAACTGACCGCATATGG
CGGCATCGACGCCCTGACCCATGCCCTCGAGGCCTATGTTAGTGTCCTCAGCACTGA
ATTCACGGAGGGATTGGCTCTTGAAGCCATCAAATTGTTATTTACTTACTTGCCGCGTG
CATACCGTTGGGGCGCAGCGGATCCGGAAGCGCGCGAGAAAGTTCACTACGCTGCGA
CTATTGCCGGCATGGCATTTGCCAACGCGTTCCTTGGTGTCTGTCATAGCATGGCCCA
CAAACTGGGTAGTACGTTCCATGTCCCACACGGTTTGGCGAATGCGCTGATGATCTCG
CATGTGATCCGCTATAACGCGACTGATGCACCGCTGAAGCAGGCGATTTTCCCCCAGT
ATAAGTATCCGCAGGCCAAAGAACGCTATGCGCAGATCGCCGACTTCTTAGAACTCGG
CGGTACCACACCTGAAGAAAAGTCGAACGCCTGATCGCCGCCATCGAGAATCTTAAG
GCCCAGCTGGAGATCCCGGCCACGATCAAAGAAGCCCTGAATTCTGAAGATCAGGCG
TTTTATGAACGCGTAGAGTCAATGGCGGAGCTGGCATTTGACGACCAGTGTACTGGCG
CGAACCCGCGCTACCCGCTCATCCAAGACCTCAAAGAACTTTACATTCTGGCCTATAT
GGGCTGTCGTCGCGATGCTGCAGCGTATCATCCCGAAGAAGCTCCGGCCAGCTAA
adhE10 ATGAGCGTCACTAATGTGACGGAACTCAATGAACTGGTCGCCCGCGTAAAAAAAGCAC
AGCGCGAATTTGCGAATTTCAGTCAGGAACAGGTCGATCGTATTTTTCGTGCGGCAGC
GCTCGCTGCCGCGGACGCTCGCATCCCGTTAGCGAAACTGGCGGTCGAAGAGTCGG
GTATGGGTATCATTGAGGACAAAGTTATCAAAAATCACTTCGCGTCGGAGTATATTTAT
AATGCTTATAAGGACGAAAAAACCTGTGGCGTTCTTTCCGAGGACTTGACATTTGGTAC
GATTACTATCGCGGAACCGATTGGGATTATCTGCGGTATTGTTCCAACTACTAACCCGA
CGTCCACCGCGATCTTTAAATCCCTGATTTCTCTGAAAACGCGTAACGCTATCATTTTC
AGCCCGCATCCGCGCGCCAAAGATGCAACTAACAAAGCTGCGGAAATCGTCTTGAAA
GCGGCAATTGCGGCCGGTGCGCCTAAAGACATTATTGGCTGGATCGATGCACCAAGC
GTTGAGATGTCAAATGCCTTGATGCATCATGATGATATCAACCTGATTCTCGCGACGG
GTGGTCCGGGTATGGTTAAGGCAGCGTATAGTAGCGGAAAACCAGCGATTGGAGTTG
GAGCGGGAAACACCCCAGTTGTCATTGACGAAACGGCGGATATTAAACGTGCGGTTG
CAAGCATTCTGATGAGTAAAACATTTGATAACGGCGTGGTCTGCGCGTCAGAGCAGTC
GGTTATCGTTGTGGACGAAATTTACACCCAGGTCCGCGAACGCTTCGCATCGCACGGC
GGTTACATTCTGCAGGGCAAAGAGCTGAAAGCGGTTCAGGACATCATCCTGAAGAATG
GCAACCTGAATGCAGCCATCGTTGGCCAAGCGGCGTACAAAATTGCTGAAATGGCAG
GCGTTTCGGTGCCTGAAACAACGAAAATTCTCATCGGAGAGGTGAAATTAATCGATGA
CTCTGAACCGTTTGCGCATGAAAAGCTGTCGCCGGTATTGGCGATGTACCGCGCGAA
GAATTTCGACGATGCGGTAGATAAAGCCGAACAGCTGGTGGAACTGGGCGGCATCGG
GCACACGTCGTGCTTGTATACTGATCAGGATAATCAGACCGCACGCATCAATTATTTTG
GCGCCAAAATGAAAACCGCACGTATTCTGATTAACACGCCGGCATCACAGGGTGGGAT
TGGGGACCTGTATAACTTCAAACTTGCGCCGAGCCTTACCCTGGGCTGTGGTAGTTGG
GGTGGAAACTCTATCAGCGAGAATGTGGGTCCGAAACATTTGATTAATACCAAAACCG
TTGCGAAGCGTGCAGAAAATATGCTCTGGCATAAACTGCCGAATTCGATTTATTTTCGT
CGTGGCTGTCTGCCCATTGCCCTGGAAGAAATCGCGACTGATGGCGCCAAACGTGCC

TTTATCGTGACGGACAGCTACTTGTTTAACAATGGTTACGTTGAAAAAGTAGTTGAAGT
GCTGAAACAACATCATATTGAGACGGATGTCTTCTTTGAAGTTGAAGCTGACCCGACG
CTCACGGTGGTGCGTAAAGGTGCAGCTCAGATGAACGCGTTTAAGCCGGATGTTATCA
TTGCCTTAGGCGGCGGTAGCCCAATGGATGCCGCTAAAATTATGTGGGTGATGTACGA
GCATCCGGAAACTCACTTTGAAGAATTGGCACTGCGTTTTATGGATATTCGTAAGCGCA
TTCATCGTTTTCCGAAAATGGGTGTCAAAGCTAAACTTGTGGCGATCACCACTACTAGC
GGTACCGGATCAGAAGTGACCCCGTTCGCCGTGGTAACGGATGATAAGACCGGTCAG
AAGTATCCCCTGGCCGACTACGCGCTTACACCGAACATGGCGATTGTTGACGCTAACC
TGGTGATGAATATGCCGAAAAGCCTGACCGCTTTCGGAGGTCTTGATGCCGTGACACA
CGCTTTAGAGGCATATGTGTCCGTTCTCGCAAACGAATACAGTGACGGCCAAGCGCTT
CAGGCCCTGAGCCTGCTGAAAGATTATCTCCCCGCTTCATATCATGAAGGGGCAACGA
ATCCGGTGGCCCGTGAACGTGTTCACAACGCAGCTACCATCGCTGGCATTGCCTTCG
CGAATGCGTTTCTGGGCGTATGTCATTCAATGGCACATAAGCTTGGTAGCGAATTCCAT
ATTCCTCACGGTCTGGCGAATGCACTGCTGATCAGCAACGTCATCCGTTATAACTCTAA
CGACAACCCGACCAAACAGACGGCCTTTTCTCAATACGACCGCCCGCAAGCTCGTCGT
CGTTATGCGGAGATCGCTACCCATCTCGGTTTGACGAAACCGGATGATCGTACTGGTG
CGAAAATCGAAAAACTGTTAGCGTGGCTTGAAGAGATCAAGGCCGATCTTGGTATCCC
AAAAAGCATTCGTGAAGCGGGCGTTGCAGAGGCGGATTTTCTGGCCAAAGTGGATAAA
CTGTCCGAAGACGCGTTTGATGACCAGTGTACAGGCGCGAACCCACGCTACCCGCTG
ATTAGTGAACTCAAAACCTTGTTGCTTGATACCTATTATGGTCGCGAATTTACCGAGCA
GACAAATACACCGGCAGCGCCCAAAGCCGAAAAAAAATCAACCAAAAATAA

adhE13 ATGGCGGTCACCAACGTGGCGGAGCTGAACGAATTGGTCGCACGTGTAAAAAAAGCC
CAGCGCGAATATGCGAACTTTACCCAAGAACAGGTGGACAAGATCTTTCGTGCTGCGG
CTCTCGCGGCGGCAGACGCGCGTATCCCGCTGGCCAAAATGGCGGTTGAGGAAAGT
GGGATGGGTATCGTTGAGGACAAAGTTATTAAGAACCACTTCGCATCAGAATACATTTA
TAACGCTTATAAGGATGAAAAGACCTGTGGTATTCTGAGTGAAGACGATACATTTGGGA
CCATTACGATTGCGGAACCGATTGGGTTAATTTGTGGTATTGTGCCAACTACCAACCC
GACCTCCACCGCAATTTTCAAAGCTTTAATCAGCCTTAAGACCCGCAATGGCATCATTT
TTTCGCCACATCCTCGTGCGAAGAACGCAACAAATAAAGCTGCCGACATTGTTCTTCA
GGCAGCCATCGCAGCAGGTGCGCCGAAAGACATTATTGGCTGGATTGATCAGCCAAC
AGTAGAACTTAGCAATCAGTTAATGCATCATCCGGATATTAACCTTATTTTGGCGACAG
GTGGCCCAGGAATGGTAAAAGCCGCCTACAGCAGCGGTAAACCTGCAATCGGGGTTG
GCGCAGGCAACACGCCAGTGGTGGTAGATGAGACCGCCGACATTAAGCGCGTTGTAG
CATCGATCCTCATGTCTAAAACCTTCGATTCGGGCGTGATCTGCGCGAGCGAGCAGAG
CGTTATTGTTGTCGATGCTATTTACGATGCGGTTCGCGAACGCTTCGCGAGTCATGGC
GGTTACTTACTCCAAGGTAAAGAGCTCAAAGCGGTCCAGGATATTATTCTTAAAAATGG
TGGGCTGAACGCCGCGATCGTGGGCCAGTCAGCCCCGAAAATTGCCGAAATGGCTGG
CATCAAGGTGCCGGCGAACACAAAGGTGCTGATTGGCGAAGTGAAACTGGTAGACGA
ATCCGAGCCGTTTGCACACGAGAAATTAAGTCCGACTCTGGCGATGTACCGCGCCAAA
GATTTTGAAGATGCAGTCGCAAAAGCAGAGAAGCTCGTCGCCATGGGCGGTATTGGC
CATACTAGTTGCCTGTATACGGATCAAGATAATCAAACCGCTCGTATTGCATATTTTGG
TGATAAAATGAAAACTGCGCGTATTCTGATTAACACGCCGGCCTCACAGGGAGGTATC
GGCGATCTTTATAATTTCAAGCTGGCTCCAAGCCTGACCCTGGGTTGCGGCTCATGGG
GCGGCAACAGCATTTCGGAGAACGTTGGGCCTAAGCATCTGATTAACAAAAAAACGGT
TGCTAAACGCGCCGAGAATATGCTCTGGCACAAATTGCCAAAATCGATCTACTTTCGC
CGTGGGAGCTTGCCGATCGCGCTGGAGGAAGTAGCTACCGATGGTGCAAAGCGTGC
GTTTATTGTCACTGACCGCTTTCTGTTTAACAATGGTTACGCAGACCAGATTACGAAAG
TCCTGAAATCTCATGGTATTGAAACAGAGGTTTTCTTCGAAGTGGAGGCGGATCCGAC
GCTCTCTATCGTACGCAAGGGTGCGGAACAAATGAACTCCTTTAAACCCGATGTGATC
ATTGCACTGGGGGGCGGTTCCCCAATGGATGCGGCCAAAATCATGTGGGTGTTATAC
GAACACCCCGAGACGCACTTTGAAGATTTAGCTTTGCGCTTTATGGATATTCGCAAACG
CATTTATAAATTTCCGAAGATGGGCGTGAAGGCGAAATGATTGCGATTACTACTACCT
CAGGTACGGGTAGTGAGGTGACGCCGTTCGCCGTGGTTACGGATGATGCAACCGGTC
AGAAATACCCTCTCGCCGATTACGCGCTGACCCCGGATATGGCGATCGTGGATGCGA
ATTTAGTTATGAACATGCCCAAAAGCTTGTGCGCGTTCGGGGGGCCTGGATGCGGTGAC

CCATGCGTTAGAAGCGTATGTGAGCGTGTTAGCGAATGAATATAGCGACGGCCAGGC
GCTGCAGGCGCTGAAATTACTGAAGGAATATCTCCCGGCCAGCTACAAAGAAGGGGC
GAAAAACCCCGTGGCCCGTGAACGCGTACATAACGCGGCCACAATTGCGGGCATTGC
GTTCGCTAACGCATTTCTTGGGGTTTGTCACAGTATGGCGCATAAATTGGGCAGCGAA
TTTCATATCCCCCACGGTTTAGCGAATGCCATGCTTATTTCCAACGTCATTCGTTATAAT
GCCAACGATAATCCCACCAAACAAACCGCATTTTCTCAGTATGACCGTCCTCAGGCCC
GCCGCCGCTATGCCGAGATTGCCGATCATTTGGGCCTGTCCGCACCGGGTGATCGCA
CAGCTCAGAAGATCGAAAAACTGCTGGCGTGGTTAGATGAGCTCAAGACAGAACTCGG
AATCCCGACGTCGATTCGCGAAGCTGGTGTGCAAGAGGCTGATTTTCTCGCAAAGTG
GATAAATTATCTGAAGATGCCTTTGACGATCAATGTACGGGTGCAAATCCACGCTACCC
GCTGATCGCCGAGCTGAAGCAGATCATGCTTGATACATTCTATGGTCGTGAGTTCAAC
GAGGCCGTCGAAGAAGAAGCGGCGGCGCCGGCAGTGGCAAAAGCGGCCGTTAAGAA
ACCTAAGAAATAA

hbdcrt2 ATGGCTGGCGAAGTGAAAACGATTACCGTAGTGGGCGCAGGTACCATGGGTCATGGC
ATTGCAGAATTGGCCGCCATCGCGGGATTTAAAGTTTACTTAGCGGATATTAATATGGA
TATTCTGAATAATGCGCTGCAACGTATCCGCTGGTCGTTGGAAAAACTTGCCGAGAAA
GGGCGCATTCGCGAATCAGTTAACACCGTTATGAGCCGTATTACTCCTATTGTTTCAGT
AAAAGACGGGGGTTATAGCGAAGAGCTCGCGAAAGCACTCAGTGAAAGCGATTTCATG
ATCGAAGCCATCCCAGAAAAACTGGAGCTGAAACAGCAACTGTTTGCTTTCGCAGATA
AACATGCTAAAGAAAGTGCAATTCTGGCGAGCAATACCAGTTCTTTACCGATTACGGAA
ATTGCAGCGGCCACAACGCGCCCGGATAAAGTCGTGGGTATGCATTTTTTTAATCCAC
CTGTGCTCATGCCCCTGGTGGAAGTGGTCAAAGGCGAAAAAACGAGTGAGGAAACTG
TCGCAGCCACTGTAGATCTGGCGAAAAAAATGGGCAAGCAAACCGTCGTAGTGAAGAA
AGATGTGCCGGGTTTTATTGTGAACCGTATCTTGGGCCGCCTCATGGAATCGGCTTGT
TTGCTGGTTGAAAAAGGTGGTTACACGGTTGTGCAGGTTGATGCGACCGCGAAATATC
TGCTGGGTCTGCCGATGGGTGTCTTCGAGTTAGCGGACTACTCAGGAATCGATGTTTT
TTACTATGTCTTCGAAGCTATGTCGCGTCGCGGTTTTAAATCCGCAAAGTGTTCGATTT
TCGAGGAAAAATTTAAGGCAGGCGAGTACGGTGTGAAATCTGGTAAAGGTATTTACTC
CTACCCCGCCCCTAACAAGTACGTGAAACCGAGCATCCCTCGCGATGAGGCTAAAGTA
GATCCTTTGCTTCTGATGGCGCTCCCGATCAATGAAGCCGCCTGGTTGCTGCGCGATG
ATGTTGCGACGAAAGAAGATATTGATAAGGCGGTGAAACTCGGCTTAGGCTGGCCAAA
AGGGGTCTTTGAATATGCAGACGAGTTCGGACTGGATAAAGTGGTTGAAGCACTGCAG
CGCATTAAACGCGATTTTGGCGTCGAGCATGCCGAACCAGATCCACTGCTGCGTCAGA
TGGTAGAGGAGGGGCGTTTGGGCCGCAAAACCGGGAAAGGTTTTTATGAGTATGGCG
AGGTCGAGGAAAGAAAATGGAAACGCTGCTGGTTCGCGTTGAAAAACCAATCGCTTG
GATTGTGCTGAATCGTCCCGATAAACTGAACGCAATTTCCCCGAAGATGATTACCGAA
CTTTCGCAGGCACTGGATGAACTCGAGGAAAACGATGACGTGCGCGTGGTTATCTTAA
CGGGAGCAGGCCGCGCGTTCTCTGCGGGGGCCGATGTGACCGCTTTTGCACAGGTG
ACTCCGATTGATATTTTACGCTTTAGTCGCAAATTCCAGGAAGTGACGCTGAAAATTCA
GTTCTACACGAAACCAGTCATCGTTGCGATTAAAGGCTATGCGTTAGGGGGTGGCTTG
GAACTGGCCATGTCAGGAGACATCCGCATTGCCTCTGAAGATGCGATGCTGGGTCAG
CCGGAGATTAATCTCGGCTTCATTCCGGGTGCGGGTGGTACCCAGCGTCTGGCACGT
CTCGCGGGTCCGGCGCGCGCCAAGGAACTGATTATGACAGGGGGATATGATTCCGGCC
CGTGATGCGGAAAAAATGGGTATTGTGAACCGCGTTGTACCACCGGAGTTACTGGAAC
AAGAAGCGCGTAGCCTGGCTCTGAAGCTGGCAGAAAAGCCACCGATCGCTCTGGCTG
CTGCCAAATACGCCATCGATTTCGGTCTGGAATCCAACATTTGGGCCGGACTGCAACT
GGAGGCGTCGCTGTTCAGCGTGCTGTTCAGCACCGAAGACGTGATTGAGGGCGTAAC
GGCATTTTTAGAAAAACGTAAACCGCGTTTTAAGGGCCGCTAA

hbdcrt3 ATGCCAAAAGTGGCGGTGATTGGTGCCGGTACCATGGGCCATGGAATCGCAGAATTAT
TTGCTATCGCGGGCTACGAAGTCGCGCTGGTCGACGTGGCCGAAGATTTTTTGAAACG
CGCACTGCAAAATATCGAGTGGTCGCTCAAAAAACTGGCAGAAAAGGCCAAATTAAA
GAAGATGTAGGCGTGATCTTAGGCCGTATCAAGCCGATCGTGAATGATGTCTGCAAGG
CAGTTGAGGGCGCTGAGCTCATGGTGGAAGCAGTGGTGGAAGATATCGAAATCAAGC
GTAAAGTGTTTGCGGAAGCGGACCGCTGTGCACCGCCGAGCGCAATTTTGGCCACCA
ACACTAGTAGCCTCCCGATTACGGAAATCGCGGAGGCGGTTAAACCCGAACGTCGTC

CGTTAGTGGTCGGAATGCATTTCTTCAACCCGCCGGTCCTGATGCCGTTGGTCGAAAT
TATTAAAGGTGCGTATACGAGTGACGAAACCGTGAAAAAACCGCTGAGTATGCATCG
AAACTTGGGAAACAAACCGTGGTGGTTAACAAAGATGTACCAGGCTTTATTGTTAATCG
CATTCTGGCTCGTCTGAACGAAGCCGCGTGTTGGATGGTCGCGCGTGGCGAAGCGAC
GATTCAGGAAGTTGACTCTGCGCTGATTTACAAAGCCGGCCTGCCGATGGGCGCCTTC
ATCCTGATGGACTACACTGGGATTGATGTGGTCTGCTTTATTGGCGACGCCATGGTTA
AACGTGGCTTTAAAACCCATCCATGCCCGGTTATCGCGGAGAAATGCAAAGAAAAAAA
GTACGGTGTCAAGAGTGGCGAAGGCTTCTACAAATACCCGGCACCGGGGAAATTCCA
GTGGCCTGAAGTACCGAAAGCGGCGGGTGATAAAGTGGATGTGACTTATTTGCTGGC
GCCGGCTATTAACGAGGCGGCTTACCTCCTGCGTGAAAGCATCGCATCACGTGAAGAT
ATTGATAAGGCGATCCGCTTAGGGCTGAATTGGCCCAAAGGTCCACTGGAGTATGCG
GATGAGTTGGGAATCGATACGGTCGTCAAGGCGCTGGAAGAGTGGAAAAATAAAACA
GGCTTCGAAGAATACGAGCCGGATCCGCTCTTGCGTGATATGGCTGCCAAAGGGAAG
CTGGGTAAAAAATCAGGCGAAGGCTTTTATACATATGTTAAAGCGGAAGAGAAAAAATT
GGAAACCCTGATCGTCCGCTATGAGCCAGGGATCGCTTGGATTATCCTGAACCGTCCC
GAACGCCTTAACGCGATTAACCCGAAAATGATTGAGGAGCTGTGGAAAGTTTTGGACG
AAATCGAGCAGATGGATTATGACAAAGTGCGTGTGGTGGTTATCACTGGTTCCGGACG
TGCCTTTTCCGCGGGCGCGGATGTCACAGGTTTTATGGGGGCCACTCCGGTGACCAT
TTTTAAAGTGTCCCGCAAACTCCAGATGCTGTATGAACGCCTCGAACTGCTGGATCGC
CCCGTAATTTGTGGCCTTAATGGTTATACGTTGGGTGGTGGCCTTGAGCTGGCCATGG
CGTGCGACTTTCGTATTGCGGCCGAAACGGCCGAACTTGGCCAGCCGGAAATCAATTT
GGGTTTCATTCCGGGTGCGGGCGGCACGCAGCGCTTGGCCCGCCTTATTGGTCGCGA
CAAAGCCAAAGAGCTGATTTTCACTGGCGACCGCATTCCGGCCCGCGAGGCCGAACG
TCTGGGCCTTGTGCACAAAGTTGTTCCACCTGACCGTCTGGAACAGGAACTGCGTGCT
TTTGCTAACAAATTGGCTGAGAAACCGCCGCTCGCCCTGGCGATGGCTAAATACGCTA
TTAACTTCGGCCTTGAAGCCCCGCAATGGGTTGGCATGATGTTGGAAGCGAGTAACTT
CGGTTTGCTGTTTAGCACGGAAGACGTGATTGAGGGCGTCAGCGCATTTTTACAGAAA
CGTAAGCCGCAGTTCAAGGGCAAATAA

hbdcrt4   ATGAAAGTTGAGGATATTAAAAAGGTGCTGGTAGTGGGGGCCGGAACCATGGGACAT
GGAATCGCTGAAATTGCGGCTATCTCCGGCTATAAAGTATATCTGTCGGATATCTCTCA
AGACATTCTGAACTCCGCCTTAGAGCGCATCCGTTGGAGCCTCTCTAAGCTGCAAGAG
CGTGGTCAGATTAAGGAGAGCATCGACACCATTATGAGTCGTATTACCACTATTGTTGG
TCTGGATAAAACGGTGAGTGACGCGGACTTTTCGATTGAAGCCAGCACCGAACGTATG
GATATCAAACGCCAGGTGTTTAGCAAACTCGATGAGTTACTGCCAAGTCACGCGATCC
TGGCCACCAATACCTCTTCACTGCCCATTACGAAGATCGCAGAAGCAACTAAACGCCC
AGACAAAGTCGTTGGCATGCACTTTTTCAATCCCCCCGTGCTTATGCAGCTGGTTGAG
GTGATGAAAGGAGATAAAACTAGTGACGAAACTGCCAAAATCACATATGACCTGGCCA
AACGCTTCGGCAAACAACCCATTATGATCAATAAAGATGTGCCTGGATACGTTGTAAAT
CGTATCTTGGGCCAAATCAATGTTGCGTCTTGTATTCTGGTTGAAAAAAAGGTTGCAGA
TTACCGCGAAGTTGATGCAGTGGCACGCTACAAGCTGGGTTTTCCTATGGGTGTGTAC
GAACTGATTGACTATACTGGTGTTGATGTGGCCTATTATGTGTCAAAAAGCCGCGAAGA
GCTGGGGATTCGCGATGATATTCCTATCTGCTCCTTGATTGAGCAGAAGTTCAAGAAC
AACGAGCTGGGTGTGAAGACGGGGAAAGGCTTCTACACTTACCCGGGACCCGGTAAG
TATGTGAAACCAGAACTGCCGAAAGAGCTGGCCGATAAGCTGAACCCGGTGCTGATC
CTGGCGGGCGCAGTGAACGAAGCCGCGCGCCTGCTGCGTGAAGGGATCGCGAGCCG
CGACGACATCGACTTAGGCGTGCGTTAGGGCTGGGTCTCCCAAAGGGCATTTTTCAA
TACGCGGATGAGCTGGGGATTGATTCTGTGGTCAAAGCACTCGAAGATCTGAAAGCCC
TGAGCGGATATTCCGTGTTTTCACCGGATCCCTTACTTACACAGATGGTCGGTGAGAA
TAAGCTGGGTATCAAGTCGGGGTCGGGATTTTATACCTATGGCAAAGTAGAAGAAAAG
AAATTGAATACGTTGATTATTCGTATCGAACCACCCCTGGCGTGGATCATTCTGAACCG
CCCTGAGCGTCTGAATGCACTGAATACCGAACTGGTGAGTGAGCTGGATAAAAGTTTG
GATGAGTTAGAAGGCCGCTCCGACGTGCGCGTTGTCATCATTACCGGCAACGGTCGC
GCGTTTAGTGCCGGAGCAGACGTATCCAGCTTCATTACGCTGCGCCCAATCGATGTGA
TCCGTCTCCGCACTTTACGCAATGTGGTTAACAAAATTGCATTGTATACCAAGCCGATC
ATTGCGGGTATTAATGGCTTCGCTCTCGGCGGCGGGTTAGAGTTAGCGATGGCTTGTG

ATATCCGTATCGCCTCAGAAGTAGCCCAACTGGGACAGCCGGAAATCAATATCGGAAT
CATCCCGGGCGCCGGTGGCACTCAGCGTCTGCCTCGCCTGGTCGGCAAAGGTAAAG
CCAAACTGATGATCTATACGGGCGACATGGTGTCGGCCGAAGATGCGTACAAAATGG
GTTTAGTGGACCTGGTGGCGCCCGCAAACCGTTTCGAGGAGGAAGTTCGTCGTGTTG
CCCTGAAAATCGCGGAAAAATCTCCGATTTCCCTGCTTGCGGCCAAACTCGCAATCGA
ATTGGGCTATGAATCTAACATCTGGACTGGCCAGACACTTGAAAGCACCCTGTTCGGT
CTGCTCTTTACCACCAAGGACGTGGAAGAGGGTGTTAAAGCGTTCCTGGAAAAACGCA
AGCCGCAGTTTAAAGGCGAATAA

hbdcrt6   ATGGAGATCCGTAAATTCACAGTCGTTGGTGCAGGGAGCATGGGTCATGGTATTGCCG
AGCTCGCTCTGATTGCAGGCTACGACGTGTGGTTGAACGACGTGGCAGAAACATCCT
GAAAAACGCAAAAGAACGTATCGGCTGGTCTCTCAAACGTCTGAGCGAAAAAGGCAGC
ATTAAAGAAGATCCTGAAAAGATTCTTAGTCGTCTTCATCTTACAGTCAGTCAAGAAGA
AGCAATGAAAGATACAGACTTTTTAATTGAGGCCGTTATCGAAGATATCAATGTGAAGA
AGAAAGTCTTTGAAAAGGCAGACAGTCTGGCCTCAAAAGACGCTATCCTCGCTAGTAA
CACGTCCAGCTTACCCATCACCGAGATTGCCACTGCTGTCAAACGCCCTGAGCGCGTA
GTTGGCATGCACTTTTTTAATCCGCCGGTATTAATGCAGCTGGTGGAGATCATCAAAG
GGGAAAAGACCTCGGAAGAAACAATGAAGCGCGCCTATGAAATGGGCAAAAAACTGG
GTCGCGATCCGATTTTAGTCAAGAAGGATGTCCCTGGATTCCTGGTGAACCGTATTTTA
TTCCGCGTAAACGATATTTCATGCTTACTGGTTGAAAAGGGTAAAGCAGATGTTCTCGA
GGTGGATGCCAGCGCGATTTATGAACTGGGCTTTCCGATGGGCGTATTTTTACTGCAA
GACTACGCAGGTGTGGATGTGGCTTACTTAGTTGGTAAAGCTATGGGCGAACGCGGG
TTCAAAGCTTACGAGTGCAAAGCATTCGAGCAGCTGTATAAAGCTAAAAACTTGGGCG
TTAAAACCGGCAAAGGCTTTTACTCTTATCCTGAGCCAAACAAATTCGTGAAACCCTCC
ATTCCCAAAGAAAAGGCAAACAAAATTCCGGGTATTATGTTAATCGCGCCGGCGATTAA
TGAAGGAGCTTACCTCGTGCGCGAAGGTATCGTGACCAAAGAAGATACGGATAAAGG
CTGCAAGCTGGGCCTGAACTGGCCTAAAGGAGTGTTCGAGTACGCCGATGAGTATGG
TCTTGACAACGTTGTGAAAACCCTGGAAGACTTGCGTAAAGAGACCGGGCTGGATTAT
TTCACCCCGGACCCATTGTTAACAAAAATGGTTAAAGAGGGTCGCCTGGGAAAGAAAG
TGGGTAAAGGATTCTACGAGTATAAATGA

Table 0.5. Codon Optimized Gene Sequences for Chapter 7.*The sequences of genes that were codon optimized by IDT's optimizer are listed.*

Gene	Sequence
EcoTHL	ATGAAAAATTGTGTCATCGTCAGTGCGGTACGTACTGCTATCGGTAGTTTTAACGGTT CACTCGCTTCCACCAGCGCCATCGACCTGGGGGCGACAGTAATTAAAGCCGCCATTG AACGTGCAAAAATCGATTCACAACACGTTGATGAAGTGATTATGGGTAACGTGTTACA AGCCGGGCTGGGGCAAAATCCGGCGCGTCAGGCACTGTTAAAAAGCGGGCTGGCAG AAACGGTGTGCGGATTCACGGTCAATAAAGTATGTGGTTCGGGTCTTAAAAGTGTGGC GCTTGCCGCCCAGGCCATTCAGGCAGGTCAGGCGCAGAGCATTGTGGCGGGGGGGTA TGGAAAATATGAGTTTAGCCCCCTACTTACTCGATGCAAAAGCACGCTCTGGTTATCG TCTTGGAGACGGACAGGTTTATGACGTAATCCTGCGCGATGGCCTGATGTGCGCCAC CCATGGTTATCATATGGGGATTACCGCCGAAAACGTGGCTAAAGAGTACGGAATTACC CGTGAAATGCAGGATGAACTGGCGCTACATTCACAGCGTAAAGCGGCAGCCGCAATT GAGTCCGGTGCTTTTACAGCCGAAATCGTCCCGGTAAATGTTGTCACTCGAAAGAAAA CCTTCGTCTTCAGTCAAGACGAATTCCCGAAAGCGAATTCAACGGCTGAAGCGTTAGG TGCATTGCGCCCGGCCTTCGATAAAGCAGGAACAGTCACCGCTGGGAACGCGTCTGG TATTAACGACGGTGCTGCCGCTCTGGTGATTATGGAAGAATCTGCGGCGCTGGCAGC AGGCCTTACCCCCCTGGCTCGCATTAAAAGTTATGCCAGCGGTGGCGTGCCCCCCGC ATTGATGGGTATGGGGCCAGTACCTGCCACGCAAAAAGCGTTACAACTGGCGGGGGCT GCAACTGGCGGATATTGATCTCATTGAGGCTAATGAAGCATTTGCTGCACAGTTCCTT GCCGTTGGGAAAAACCTGGGCTTTGATTCTGAGAAAGTGAATGTCAACGGCGGGGGCC ATCGCGCTCGGGCATCCTATCGGTGCCAGTGGTGCTCGTATTCTGGTCACACTATTAC ATGCCATGCAGGCACGCGATAAAACGCTGGGGCTGGCAACACTGTGCATTGGCGGC GGTCAGGGAATTGCGATGGTGATTGAACGGTTGAATTAA
CacTHL	ATGAAGGAAGTGGTGATAGCTAGTGCAGTGCGGACCGCTATTGGGAGTTACGGCAAG TCATTGAAGGATGTCCCTGCTGTTGATTTGGGAGCCACCGCGATTAAAGAGGCCGTAA AGAAAGCTGGCATAAAGCCCGAGGATGTCAATGAAGTTATCCTGGGGAACGTTTTGCA AGCTGGCTTGGGGCAAAATCCGGCCCGGCAAGCATCTTTTAAAGCCGGCCTTCCAGT AGAAATACCCGCTATGACGATCAACAAGGTATGCGGTAGCGGACTTAGAACAGTGTC GCTTGCGGCTCAGATAATTAAGGCAGGGGACGCTGACGTTATCATTGCGGGTGGTAT GGAGAACATGAGTCGTGCGCCCTACCTGGCGAACAATGCTAGATGGGGTTATCGCAT GGGGAACGCGAAGTTCGTCGATGAAATGATAACTGACGGCCTTTGGGACGCATTTAA TGACTACCACATGGGAATCACCGCTGAGAACATTGCCGAACGCTGGAATATATCGAGA GAAGAGCAGGACGAATTTGCCCTTGCCTCACAGAAAAAGGCGGAAGAGGCCATTAAA TCTGGACAATTCAAAGATGAAATCGTCCCAGTCGTGATAAAGGGCAGAAAAGGGGAAA CTGTTGTGGACACGGATGAGCACCCCCGGTTCGGGTCAACAATAGAGGGCTTGGCAA AACTGAAACCCGCGTTCAAGAAAGATGGTACAGTCACCGCGGGTAACGCATCGGGGT TGAATGATTGCGCGGCGGTATTGGTGATTATGTCTGCTGAAAAGGCTAAAGAATTAGG AGTAAAACCTTTGGCCAAAATTGTCAGCTATGGGAGTGCTGGAGTAGACCCCGCGAT CATGGGATATGGCCCGTTCTACGCCACAAAAGCTGCTATTGAGAAAGCTGGGTGGAC CGTTGATGAGCTGGACTTGATTGAGTCAAATGAAGCATTCGCCGCTCAGTCGTTGGC GGTGGCTAAGGATCTTAAATTTGATATGAACAAGGTCAATGTAAACGGAGGCGCGATC GCATTAGGACATCCTATAGGTGCAAGCGGAGCACGCATTCTGGTTACTTTAGTCCACG CTATGCAAAAGCGGGACGCTAAGAAGGCCTGGCTACACTTTGTATCGGCGGAGGCC AGGGCACTGCCATTTTGTTAGAAAAATGCTAA
CklTHL1	ATGCGTGAAGTAGTGATAGTATCTGCCGTTCGCACGGCTATAGGATCATTCGGGGGT ACTTTGAAGGATGTATCTGCAGTAGATTTGGGTGCTATTGTAATAAAGGAAGCTGTAAA GCGGGCGGGTATTAAGCCCGAGCAAGTGGATGAGGTAATTTTTGGTAACGTGATACA GGCGGGTGTAGGACAGTCATTAGCAAGACAGTCAGCCGTGTACGCCGGCTTGCCCG TCGAGGTACCTGCGTTTACAGTGAATAAGCTGTGCGGTAGCGGACTTCGCACAGTAT CTCTTGCTGCCTCCTTGATCTCGAACGGTGATGCGGACACAATAGTCGTTGGCGGCA GTGAAAATATGTCTGCGAGCCCTTATTTAATACCCAAGGCTCGGTTCGGTTACCGTAT GGGCGAAGCCAAAATCTATGATGCAATGCTGCACGATGGTTTGATAGATTCGTTCAAC AACTACCACATGGGAATTACCGCCGAGAATATAGCGGAGAAATGGGGTATTACGAGA

GAGGATCAGGACAAATTCGCTTTAGCTAGTCAGCAGAAGGCCGAAGCAGCGATCAAA
GCTGGCAAATTCAAAGACGAAATCGTACCTGTAACGGTCAAAATGAAAAAAAAGAGG
TCGTGTTCGACACCGACGAGGATCCGCGCTTTGGGACTACAATTGAAACTTTAGCGAA
ATTGAAGCCTGCTTTTAAACGGGATGGGACTGGTACCGTCACGGCAGGAAACAGTTC
TGGGATCAACGATTCTAGTGCCGCACTTATCCTGATGTCGGCTGATAAGGCTAAGGAA
CTTGGGGTTAAACCGATGGCAAAATATGTAGATTTTGCCTCGGCAGGGCTTGATCCTG
CAATTATGGGTTATGGTCCATATTATGCCACAAAGAAAGTATTGGCTAAAACTAATCTT
ACGATTAAAGATTTTGATTTGATAGAGGCTAACGAGGCTTTCGCTGCTCAATCGATTGC
AGTCGCGCGTGACTTGGAGTTTGACATGTCGAAGGTTAATGTGAACGGTGGGGCCAT
AGCTCTGGGGCATCCTGTGGGATGTAGTGGGGCACGTATCCTTGTTACCCTGCTGCA
CGAAATGCAGAAACGCGACGCAAAGAAGGGCTTGGCTACCTTATGTATTGGGGGAGG
TCAAGGAACAGCGGTCGTAGTGGAGCGCTAA

ReuTHLa	ATGACAGACGTCGTAATCGTTTCGGCAGCAAGAACAGCAGTTGGGAAATTTGGAGGG AGTTTAGCTAAAATTCCCGCTCCCGAGCTGGGGGCCGTAGTGATCAAAGCGGCGTTA GAGCGCGCTGGTGTGAAACCCGAGCAGGTTAGTGAAGTAATCATGGGACAGGTACTG ACGGCTGGTAGTGGACAAAATCCAGCTAGACAAGCTGCAATCAAGGCAGGACTTCCT GCGATGGTGCCTGCGATGACCATTAACAAGGTGTGCGGGTTCAGGATTGAAGGCGGTC ATGTTAGCAGCCAACGCGATAATGGCCGGAGACGCCGAGATAGTAGTGGCTGGAGG CCAAGAGAATATGTCTGCGGCCCCCCACGTCCTTCCAGGTAGTAGAGATGGTTTTCG GATGGGAGACGCCAAGTTAGTAGACACTATGATTGTAGACGGTTTATGGGACGTATAC AACCAATATCACATGGGAATAACGGCTGAGAACGTCGCTAAGGAATATGGGATAACTC GGGAAGCCCAGGACGAATTTGCGGTAGGATCGCAAAATAAAGCCGAGGCTGCTCAAA AGGCGGGGAAGTTTGATGAGGAAATTGTTCCTGTTTTAATTCCGCAGCGGAAAGGAG ACCCAGTAGCATTCAAAACGGACGAGTTTGTCCGCCAGGGCGCTACCTTGGACTCAA TGTCGGGTCTTAAACCTGCGTTCGATAAAGCAGGCACTGTTACAGCCGCCAACGCCA GCGGCTTAAATGATGGCGCGGCTGCAGTCGTCGTAATGTCAGCAGCCAAGGCGAAAG AGCTTGGCTTGACACCATTAGCTACGATTAAATCGTACGCGAATGCTGGTGTAGACCC CAAAGTCATGGGAATGGGGCCTGTACCCGCGTCGAAGAGAGCATTGTCCCGGGCGG AATGGACGCCACAAGACCTGGACCTGATGGAGATCAATGAAGCTTTCGCTGCACAGG CTTTGGCGGTTCATCAACAGATGGGATGGGATACTTCCAAAGTCAACGTGAACGGAG GGGCTATCGCAATAGGTCATCCTATCGGCGCGAGCGGCTGCCGGATACTTGTCACTT TATTGCATGAAATGAAGCGGAGAGATGCCAAGAAGGGATTGGCATCTCTGTGTATAGG AGGTGGCATGGGTGTAGCTTTGGCTGTAGAACGTAAGTAA
ReuTHLb	ATGACACGTGAAGTTGTCGTAGTAAGTGGGGTTCGTACGGCCATAGGTACTTTCGGA GGTTCTCTGAAGGATGTTGCCCCCGCCGAGTTAGGTGCATTAGTAGTACGTGAGGCA TTAGCGCGGGCCCAAGTGTCGGGTGATGACGTAGGGCATGTGGTTTTCGGCAACGTC ATACAGACTGAGCCACGTGATATGTACTTGGGTCGGGTAGCCGCCGTGAACGGCGG GGTGACAATTAACGCTCCGGCCTTGACGGTCAATCGGCTTTGCGGCAGCGGGCTTCA AGCTATTGTCAGTGCAGCCCAGACCATTCTTTTGGGTGATACCGATGTCGCAATCGGC GGAGGAGCAGAATCAATGTCGCGCGCGCCATATTTAGCGCCAGCAGCGAGATGGGG GGCCCGGATGGGTGATGCAGGGTTAGTAGATATGATGTTAGGAGCGTTGCATGACCC ATTTCATCGTATACACATGGGAGTGACAGCCGAAAACGTCGCTAAAGAATACGACATC TCGCGCGCGCAACAAGATGAGGCAGCATTGGAGAGTCACAGACGTGCCTCAGCAGC TATAAAAGCTGGGTATTTCAAGGACCAGATCGTACCTGTTGTATCCAAGGGCCGGAAA GGTGACGTTACTTTTGACACTGACGAACACGTCCGCCACGATGCTACCATTGATGATA TGACGAAATTACGTCCTGTGTTTGTAAAGGAAAACGGAACTGTTACCGCTGGGAACGC CTCAGGGCTGAACGACGCGGCCGCTGCCGTTGTAATGATGGAACGGGCCGAGGCGG AACGTCGTGGTTTGAAACCGCTGGCACGGTTGGTAAGCTATGGCCACGCTGGCGTAG ATCCAAAGGCAATGGGTATCGGACCTGTTCCAGCAACTAAAATTGCTCTTGAACGCGC TGGTCTTCAAGTCAGTGATTTAGATGTAATCGAAGCAAATGAGGCGTTCGCCGCACAA GCTTGTGCCGTAACCAAGGCGCTGGGGTTGGATCCAGCAAAGGTGAACCCCAATGG GAGTGGCATATCATTGGGGCACCCTATAGGTGCGACAGGCGCGTTGATTACTGTCAA GGCGCTGCATGAGTTAAATCGCGTACAGGGCCGTTACGCGCTTGTCACAATGTGTAT AGGAGGGGGCCAGGGGATTGCCGCCATTTTCGAACGCATCTAA

CklTHL2	ATGAAAGATGCAGTTATTGTAAGTGCAGTAAGAACAGCTATAGGGAGTTTTGGTGGAA CTTTAAAAGATATTTCTGCTGTAGATTTGGGGGCAATAGTTATAAAAGAGGCTGTAAAA AGAGCAGGTATAAAACCAGAACAAGTAGATGAAGTTATATTTGGAAATGTAATACAGG CAGGTCTTGGACAAAGTCCAGCGAGGCAAGCTGCTGTAAAAGCAGGCATTCCTGTAG AAGTACCAGCGTTTACACTAAATAAGGTTTGCGGTTCAGGACTTAGATCAGTAAGTTTG GCAGCTCAGCTCATAAAAATTGGAGATGATGATATTGTTGTAGTTGGTGGAACAGAAA ACATGTCCGCTGCACCATATCTACTTCCAAAGGCCAGATGGGGGACATAGAATGGGAG AGGGAAAATTAGTTGATGCCATGATAAAAGATGGACTTTGGGAAGCATTTAACAATTAC CACATGGGAATTACAGCTGAAAACATAGCAGAAAATGGGGAATAACAAGAGATATGC AGGATGAATTTGCATTAGCATCCCAACAGAAGGCAGAAGCAGCCATAAAGGCAGGAA AATTTAAAGATGAAATAGTTCCAGTAACCGTTAAGCAGAAAAAGAAAGAAATAATTTTT GATACTGATGAATTCCCTAGATTTGGGACAACTATAGAAGCATTAGCAAAATTGAAACC ATCATTCAAAAAAGATGGAACAGTTACAGCAGGTAATGCTTCGGGAATAAATGATGCA GCAGCAGCTTTAGTTGTAATGAGTGCAGATAAGGCAAAAGAACTTGGAATTAAGCCTC TTGCAAAGATTGTTTCCTATGGAAGTAAAGGATTAGACCCAACCATAATGGGATACGG ACCTTTCTATGCAACAAAGTTGGCACTTGAAAAAGCTAACTTGTCAATTGCAGATTTAG ACTTAATAGAAGCAAATGAAGCATTCGCTTCACAAAGTTTAGCAGTAGCAAAAGATTTA GAATTTGATATGAGCAAAGTAAATGTAAATGGAGGAGCAATAGCTCTTGGACATCCAG TTGGCTGCTCTGGTGCAAGAATACTCGTTACATTACTTTATGAAATGCAGAGAAGAGAT GCGAAAAAGGGACTTGCAACATTATGTATAGGGGGAGGAATGGGAACTGCACTAATA GTTGAAAGATAA
CacTHLm	ATGAAAGAAGTTGTAATAGCTAGTGCAGTAAGAACAGCGATTGGATCTTATGGAAAGT CTCTTAAGGATGTACCAGCAGTAGATTTAGGAGCTACAGCTATAAAGGAAGCAGTTAA AAAAGCAGGAATAAAACCAGAGGATGTTAATGAAGTCATTTTAGGAAATGTTCTTCAAG CAGGTTTAGGACAGAATCCAGCAAGACAGGCATCTTTTAAAGCAGGATTACCAGTTGA AATTCCAGCTATGACTATTAATAAGGTTTGTGGTTCAGGACTTAGAACAGTTAGCTTAG CAGCACAAATTATAAAAGCAGGAGATGCTGACGTAATAATAGCAGGTGGTATGGAAAA TATGTCTAGAGCTCCTTACTTAGCGAATAACGCTAGATGGGGGATATGGAATGGGAAAC GCTAAATTTGTTGATGAAATGATCACTGACGGATTGTGGGATGCATTTAATGATTACAA TATGGGAATAACAGCAGAAAACATAGCTGAGAGATGGAACATTTCAAGAGAAGAACAA GATGAGTTTGCTCTTGCATCACAAAAAAAAGCTGAAGAAGCTATAAAATCAGGTCAATT TAAAGATGAAATAGTTCCTGTAGTAATTAAAGGCAGAAAGGGAGAAACTGTAGTTGATA CAGATGAGCACCCTAGATTTGTATCAACTATAGAAGGACTTGCAAAATTAAAACCTGCC TTCAAAAAAGATGGAACAGTTACAGCTGGTAATGCATCAGGATTAAATGACTGTGCAG CAGTACTTGTAATCATGAGTGCAGAAAAAGCTAAAGAGCTTGGAGTAAAACCACTTGC TAAGATAGTTTCTTATGGTTCAGCAGGAGTTGACCCAGCAATAATGGGATATGGACCT TTCTATGCAACAAAAGCAGCTATTGAAAAAGCAGGTTGGACAGTTGATGAATTAGATTT AATAGAATCAAATGAAGCTTTTGCAGCTCAAAGTTTAGCAGTAGCAAAAGATTTAAAAT TTGATATGAATAAAGTAAATGTAAATGGAGGAGCTATTGCCCTTGGTCATCCAATTGGA GCATCAGGTGCAAGAATACTCGTTACTCTTGTACACGCAATGCAAAAAAGAGATGCAA AAAAAGGCTTAGCAACTTTATGTATAGGTGGCGGACAAGGAACAGCAATATTGCTAGA AAAGTGCTAG
CbeHBD	ATGAAAAAGATTTTTGTGTTGGGCGCGGGCACCATGGGTGCGGGTATCGTGCAGGCG TTCGCGCAGAAAGGTTGCGAAGTGATCGTGCGCGACATTAAGGAAGAATTTGTGGAC CGCGGCATTGCGGGCATCACCAAAGGCCTGGAAAAGCAGGTGGCGAAAGGCAAAAT GAGCGAAGAAGATAAAGAAGCGATTTTAAGCCGCATCAGCGGCACCACCGATATGAA ACTGGCGGCGGACTGCGATCTGGTGGTGGAAGCGGCGATCGAAAATATGAAAATCAA GAAGGAAATCTTCGCGGAACTGGATGGCATCTGCAAGCCGGAAGCTATCCTGGCGAG CAATACCAGCAGCCTGAGCATCACCGAAGTGGCGAGCGCGACCAAGCGCCCGGATA AAGTGATCGGCATGCATTTCTTTAACCCGGCGCCGGTGATGAAGTTGGTGGAAATCAT CAAAGGCATTGCGACCAGCCAGGAAACCTTTGATGCGGTGAAGGAACTGAGCGTGGC GATCGGCAAAGAACCGGTGGAAGTGGCGGAAGCGCCGGGCTTCGTGGTGAATCGCA TTCTGATCCCGATGATCAATGAAGCGAGCTTTATCTTACAGGAAGGCATTGCGAGCGT GGAAGATATCGATACCGCGATGAAATATGGTGCGAATCATCCGATGGGCCCGCTGGC GCTGGGCGATTTGATCGGCCTGGACGTGTGTCTGGCGATCATGGATGTGCTGTTCAC

	CGAAACCGGTGATAATAAGTACCGCGCGTCATCAATTCTGCGCAAATATGTGCGCGC GGGCTGGTTGGGCCGCAAAAGCGGCAAAGGCTTCTATGATTACAGCAAATAA
CacHBD	ATGAAAAAAGTATGTGTCATCGGTGCCGGCACGATGGGCTCGGGGATTGCTCAAGCC TTCGCTGCAAAGGGATTCGAAGTTGTGCTGCGTGATATCAAGGACGAATTTGTGGATC GCGGCCTGGATTTCATCAACAAAAATCTCAGCAAACTGGTGAAGAAAGGCAAAATTGA GGAAGCCACTAAAGTGGAAATTCTGACCCGTATTTCCGGCACGGTTGACCTGAATATG GCGGCCGATTGTGACCTGGTTATTGAAGCGGCGGTCGAACGCATGGATATCAAGAAA CAAATCTTTGCCGATCTTGATAACATTTGCAAACCGGAGACTATCCTCGCCTCAAATAC AAGCAGTTTAAGTATTACCGAAGTGGCAAGCGCTACAAAACGGCCCGATAAAGTGATT GGAATGCATTTTTTCAACCCAGCCCCGGTTATGAAACTGGTTGAAGTGATTCGCGGCA TCGCTACCTCCCAAGAAACCTTTGATGCAGTTAAAGAAACCTCGATCGCCATTGGTAA AGATCCAGTGGAGGTAGCCGAAGCGCCGGGCTTCGTGGTTAATCGGATCTTAATTCC GATGATTAACGAAGCTGTTGGCATTCTGGCCGAAGGCATTGCGTCCGTGGAAGACAT CGACAAAGCAATGAAATTGGGTGCAAATCACCCTATGGGTCCACTCGAACTTGGCGAT TTTATCGGTCTTGATATTTGCCTGGCGATCATGGACGTGCTGTATTCAGAGACAGGCG ATAGTAAATACCGCCCGCACACGCTGCTGAAAAAATATGTTCGGGCTGGCTGGCTGG GGCGTAAATCTGGTAAGGGTTTTTTACGATTATTCCAAATAA
CpaHBD	ATGAAAAAGATCTTTGTGTTGGGCGCCGGCACCATGGGTGCCGGTATTGTGCAGGCC TTTGCACAGAAAGGGTGCGAAGTGATTGTGCGCGATATCAAAGAAGAATTTGTAGATC GCGGAATCGCTGGTATTACGAAAGGGTTAGAAAAACAAGTGGCTAAAGGGAAAATGA GCGAGGAGGATAAAGAGGCCATTCTTTCGCGCATTAGCGGCACCACCGATATGAAAT TAGCTGCGGATTGTGATCTGGTGGTTGAAGCAGCAATTGAAAACATGAAAATCAAAAA AGAAATTTTTGCCGAGCTGGATGGCATTTGTAAACCGGAAGCCATTTTAGCCTCAAAT ACCTCTAGCCTGAGTATCACCGAAGTAGCCAGCGCGACCAAACGCCCCGATAAAGTT ATTGGAATGCATTTCTTCAACCCTGCACCAGTGATGAAACTGGTGGAAATTATTAAGG GAATTGCAACCAGTCAAGAAACGTTTGATGCGGTTAAAGAACTGTCGGTCGCTATTGG CAAAGAGCCAGTGGAAGTCGCCGAAGCCCCGGGCTTTGTGGTCAATCGGATTCTGAT CCCGATGATTAACGAAGCCAGCTTTATCTTGCAGGAAGGAATTGCGAGCGTGGAAGA TATCGATACGGCGATGAAATACGGGGCAAATCACCCGATGGGCCCGCTGGCTTTGGG GGACCTGATTGGCCTGGACGTTTGCCTGGCGATTATGGACGTGTTGTTTACTGAAACC GGCGACAACAAGTATCGTGCGAGTTCAATCCTCCGTAAATATGTGCGGGCCGGGTGG CTCGGTCGCAAATCGGGCAAAGGCTTTTATGACTACAGCAAATAA
CsaHBD	ATGAAGATTTTCGTGTTGGGAGCGGGGACAATGGGGGCTGGGATCGTCCAGATTTTT GCAGAGGCCGGTTATCAGGTGATCATGCGTGATATCGAAGAGAGTTTCGTCCAGAAG GGTATCACAAATATTACTAAAAACTTAGACAAAGCCGTTAAAAAAGAAAAAATCACGGA GGAAAGCAAAAACGAAGTGCTGGGACGCATCATCGCCACCACGGACATTAACCTTGC AAAAGACGCAGATTTAGTTATCGAAGCAGCCATTGAAAACATGAATATTAAAAAAAAGA TCTTTGCGGAGCTTGACGACGTTTGTAAACCCGAAACTATTCTGGCGACAAACACGTC ATCCTTAAGTATCACCGACGTGGCATCCGCGACTAAGCGTCCTGACAAGGTTATTGGG ATGCACTTTTTTAATCCTGTTCCAGTCATGAAACTGGTAGAAGTAATCACCGGTATGGC GACGTCGGCGGAAACGAAAGATACCGTTATTGAAATTACCAAGAAGGTAGGTAAGGAT CCGGTAGAAGTGAAAGAAGCACCGGGCTTTGTAGTGAATCGCATTTTAATCCCGATGA TCAATGAAGCGGTAGGTATCCTGGCGGATAATGTCGCTACCGCCGAAGATATTGATAT CGCAATGAAACTGGGCGCGAACCACCCGATGGGTCCGCTGGCCCTGGCCGATCTGA TTGGGAACGATGTGTGTCTGGCCATCATGGAAATTCTGTACATTGAATTTGGGGATCC TAAATATCGGCCGAATCCAATGCTGCGGAAAATGGTGCGCGCAGGTTATCTGGGCCG TAAAACGGGCAAGGGCTTTTATGATTATTCCAAGTAA
CklHBD1	ATGAGCATCAAATCTGTGGCCGTACTGGGCTCGGGGACGATGAGCCGTGGTATTGTT CAGGCTTTCGCAGAAGCGGGTATCGATGTGATCATCCGCGGTCGCACGGAAGGCAGT ATCGGGAAAGGGCTTGCTGCTGTTAAAAAGGCGTATGATAAGAAAGTCTCAAAAGGTA AAATTAGCCAAGAAGACGCAGACAAAATCGTGGGCCGTGTGAGTACAACCACTGAGC TCGAAAAACTGGCTGATTGTGATCTCATCATCGAAGCGGCCTCTGAGGACATGAACAT TAAAAAAGACTACTTCGGCAAGCTGGAAGAGATCTGCAAACCTGAAACAATTTTTGCG ACGAACACATCTTCGTTGTCCATCACGGAAGTCGCGACAGCGACTAAACGCCCGGAT AAATTCATCGGTATGCATTTTTTCAATCCGGCAAACGTTATGAAATTAGTTGAGATTATC

CGCGGGATGAATACGTCCCAGGAGACGTTTGACATCATCAAAGAAGCCAGCATCAAA
ATTGGCAAAACCCCTGTGGAAGTGGCGGAAGCGCCGGGTTTTGTGGTTAACAAAATC
CTGGTGCCAATGATCAACGAAGCCGTTGGCATCCTGGCCGAAGGGATTGCATCAGCG
GAAGACATTGACACTGCAATGAAACTGGGCGCCAACCATCCTATGGGGCCGCTTGCC
CTCGGAGACTTAATTGGGTTAGACGTGGTCTTAGCTGTGATGGATGTGCTGTATTCGG
AAACCGGCGACTCTAAATACCGTGCGCATACTCTGCTGCGCAAGTATGTCCGTGCAG
GTTGGCTGGGCCGCAAAAGCGGTAAAGGTTTTTTCGCCTACTAA

CklHBD2	ATGGATATCAAAAATGTGGCCGTACTCGGCACGGGCACTATGGGTAACGGCATCGTC

CklHBD2   ATGGATATCAAAAATGTGGCCGTACTCGGCACGGGCACTATGGGTAACGGCATCGTC
CAGCTGTGCGCTGAGAGCGGTCTTAATGTAAATATGTTTGGTCGGACCGATGCTAGC
CTCGAACGCGGATTTACAAGTATCAAAACGTCCCTGAAAAACCTGGAGGAAAAAGGGA
AAATTAAAAACGAATATTTCTAAAGAAATTCTGAAGCGTATCAAAGGCGTAAAAACAATT
GAAGAAGCAGTCGAAGGCGTGGACTTCGTGATTGAATGTATTGCGGAAGACCTGGAA
CTGAAACAAGAAGTCTTTAGCAAGCTGGACGAGATCTGTGCTCCCGAAGTGATCTTAG
CGAGCAATACCAGTGGCCTGTCGCCGACCGACATCGCTATCAACACGAAACACCCGG
AGCGGGTTGTAATTGCGCACTTTTGGAACCCGCCACAGTTTATTCCGCTGGTAGAGGT
TGTGCCGGGAAAACATACTGATAGTAAAACCGTGGACATCACCATGGATTGGATCGAA
CATATCGGTAAAAAAGGCGTGAAAATGCGCAAAGAGTGCCTGGGGTTTATCGGCAAC
CGTCTGCAACTGGCCCTTCTGCGTGAGGCACTTTATATCGTTGAACAAGGTTTCGCCA
CGGCGGAGGAAGTTGATAAGGCAATTGAGTATGGGCATGGCCGGCGTCTCCCTGTGA
CGGGCCCGATCTGTTCCGCGGATCTGGGCGGTCTGGATATTTTCAATAACATCAGTTC
GTATTTGTTTAAAGATTTATGTAACGATACTGAACCAAGCAAGCTTTTGAAATCGAAAG
TCGACGGCGGTAATCTGGGCTCTAAAACCGGTAAAGGTTTCTATAACTGGACACCCGA
GTTCTTACAAAAAAAGCAGAATGAACGTATTCAGCTGCTGATGGACTTCCTGGAAAAA
GACAAAAACGATAAAAGCATTGAACGCAACATTTAA

CacCRT   ATGGAACTTAATAACGTAATCTTAGAAAAGGAAGGTAAAGTGGCGGTGGTGACAATCA
ATCGCCCGAAAGCGCTGAACGCTCTGAACAGCGATACCCTCAAAGAAATGGATTATGT
GATTGGTGAAATCGAAAACGATTCAGAAGTGTTAGCGGTGATCCTGACCGGCGCGGG
CGAAAAAAGCTTTGTGGCGGGCGCGGATATCAGCGAGATGAAGGAAATGAACACAAT
CGAAGGTCGCAAATTCGGAATTTTGGGCAACAAAGTATTTCGCCGCCTGGAATTATTA
GAAAAGCCGGTGATTGCGGCGGTGAACGGTTTTGCGCTGGGCGGAGGCTGTGAAAT
TGCGATGAGCTGCGATATTCGCATTGCGAGCTCAAATGCGCGCTTTGGTCAGCCGGA
AGTGGGTCTAGGCATTACCCCGGGTTTTGGTGGTACCCAGCGCTTAAGCCGCCTGGT
GGGCATGGGAATGGCGAAGCAATTGATTTTTACCGCGCAGAACATTAAGGCGGATGA
AGCGCTGCGCATCGGCTTGGTGAACAAGGTGGTGGAACCGAGCGAACTGATGAACA
CCGCGAAAGAAATCGCGAATAAAATCGTATCAAACGCGCCGGTGGCGGTGAAGCTGT
CAAAACAAGCGATCAACCGCGGCATGCAATGCGATATCGATACCGCGCTGGCGTTTG
AAAGCGAAGCGTTTGGCGAATGTTTTAGCACCGAGGATCAGAAGGATGCGATGACCG
CGTTCATTGAGAAACGCAAAATCGAAGGATTCAAAAACCGCTGA

PpuCRT   ATGACAACCCCGAGCAGCCCTCTGTTAAGCAAAGTTGAGGCTGGCGTAGCGTGGATT
ACCTTGAACCGCCCAGAACAGCGCAACGCCCTGGATATCCCAACCTTAAAACAACTG
CATGCGTTATTAGATAGCCACGCGGATGATCCAGCGGTACGCGTGGTGGTGCTGACC
GGCAGCGGCCGCAGCTTTTGCGCTGGCGCGGATCTGGCGGAGTGGGCTGCGGCGG
AGGCTGCGGGCACCCTGGAGAGCTACGGCTGGACCGAGACAGCGCACGCGCTGATG
TTGCGCCTGCATAGCTTGGATAAGCCAACCATTGCGGCGATTAACGGCACCGCGGTG
GGCGGGGGCATGGATCTCAGCCTGTGCTGCGATCTGCGCATTGCGGCGGCGAGCGC
CCGCTTTAAAGCGGGCTATACCAGCATGGGCTATAGCCCAGACGCGGGCGCGAGCT
GGCATCTGCCTCGGCTGATTGGCAGCGAACAGGCGAAACGCTTGTTATTTTTGGACG
AGCTGTGGGGCGCGGAACACGCGCTGGCCGCTGGGCTGGTTAGCGAGGTTTGCGC
GGATGAACAACTGCCAGCGGTGGCCGCGGAATTAGCGGGGCGCCTGGCGAATGGCC
CGACTTTTGCGTACGCCCAGACCAAACAGCTGATTCGCGATGGCGCGCGGCGCACCT
TAGCGGAACAGCTGGAAGCTGAACGCCATGCGGGCCTGCTGTGCGGCCGCAGCCAG
GACGGCGCGGAAGCGCTGCAAGCGAGCGTAGAGCGCCGCGCGCCACGGTTCACCG
GCCAGTGA

CbeCRT   ATGGAGTTAAAGAATGTAATCCTTGAAAAGGAGGGCCACCTTGCTATAGTTACGATCA
ATCGCCCGAAGGCATTAAATGCACTGAACTCAGAAACCCTTAAAGACTTGAATGTTGT

TCTGGACGATCTTGAAGCCGACAACAATGTTTACGCCGTAATCGTCACAGGAGCAGG
CGAAAAGTCGTTTGTAGCTGGCGCGGACATCGCAGAGATGAAAGACTTAAATGAAGA
GCAAGGAAAAGAGTTCGGGGATACTGGGCAACAATGTCTTCAGAAGACTTGAAAAATTA
GATAAGCCCGTAATTGCAGCTGTGAGCGGTTTTGCATTAGGTGGGGGGCTGCGAGCTG
GCTATGAGCTGCGACATACGCATAGCATCGGTTAAGGCCAAATTCGGTCAACCCGAG
GTTGGATTGGGCATAACGCCGGGATTCGGCGGTACTCAGCGGTTAGCAAGAATTGTT
GGGCCGGGGAAAGCTAAAGAACTTATATACACTTGTGACATCATAAACGCCGAAGAAG
CCTACCGGATTGGGTTAGTTAATAAGGTAGTTGAGTTGGAGAAGCTGATGGAAGAGG
CAAAAGCGATGGCAAACAAGATTGCAGCCAATGCTCCCAAAGCTGTCGCATATTGCAA
GGACGCTATTGATCGGGGGATGCAAGTTGACATTGACGCCGCTATATTGATAGAAGC
GGAAGACTTTGGGAAATGTTTCGCAACGGAAGATCAAACGGAAGGAATGACAGCATT
CTTGGAAAGACGCACCGAAAAGAACTTCCAGAATAAGTAA

CklCRT　ATGGAGTTTAAGAATATAATTCTGGAGAAAGACGGGAACGTCGCTTCCATAACATTAAA
TCGCCCGAAAGCCCTGAATGCCTTAAATGCTGCTACGCTGAAGGAAATCGACGCAGC
AATCAATGACATCGCTGAAGACGACAATGTTTATGCCGTGATAATCACAGGTTCGGGG
AAAGCATTCGTCGCGGGAGCCGATATCGCAGAAATGAAGGACTTAACGGCCGTAGAG
GGTCGTAAATTTTCGGTGTTGGGCAATAAGATATTTCGCAAGCTGGAGACCTTGGAAA
AGCCAGTGATTGCAGCTATTAACGGATTCGCACTGGGTGGAGGATGCGAGTTGTCCC
TTTCATGCGATATACGCATAGCGTCGAGTAAGGCGAAATTCGGGCAACCCGAGGTTG
GCTTAGGGATCACCCCAGGCTTCGGAGGGACTCAGCGCCTGGCCCGTGCTATTGGC
GTGGGAATGGCAAAAGAACTGATTTACACCGGTAAGGTCATAAACGCCGAAGAGGCA
CTTCGGGTCGGACTGGTAAATAAAGTGGTCGAGCCAGATAAGTTATTAGAAGAAGCAA
AGTCTCTGGTGGACGCGATCATTGTTAATGCTCCAATAGCCGTACGGATGTGCAAAGC
TGCCATAAACCAAGGATTGCAGTGTGATATTGATACCGCAGTTGCATACGAAGCAGAG
GTTTTCGGGGAATGTTTTGCTACGGAAGATCGTGTCGAGGGCATGACGGCTTTCGTG
GAGAAGCGTGATAAGGCTTTTAAGAATAAGTAA

CpaCRT　ATGGAGCTGAAAAACGTCATACTGGACAAGGAAGGAAAGATCGCTGTGGTTACCATTA
ACCGTCCGAATGCTCTTAATGCACTTAATTCCGAGACACTTAAAGAATTGGATTACGTC
ATCGATGAAATAGAAAACGATTCGAACGTCTTTGCCGTTATTCTTACAGGAGCTGGTG
AGAAATCATTTGTCGCTGGAGCGGACATCGCCGAGATGAAGGACATGAACACCATCG
AGGGTCGGAAATTTGGTATTTTAGGAAATCGTGTATTTCGTCGTATTGAACTGCTGGAA
AAGCCAGTGATTGCGGCGGTCAATGGGTTTGCCCTGGGTGGCGGGTGTGAGCTGAG
CATGTCATGCGATATTAGAATCGCTTCCTCGAACGCCCGGTTTGGACAGCCTGAGGTC
GGATTGGGCATTACTCCGGGGGTTTGGAGGTACACAGCGTCTGGCTCGTTTGGTTGGC
ATGGGTATGGCGAAGCAGATTATTTTCACTGCCAAGAATATCAAAGCGGATGAAGCAC
TGAGAATTGGGTTGGTCAACAAGGTGGTGGAGCCGGGAGAATTGATGGATACTGCAA
AGGATATTGCAAACACAATTGCATCCAAGGCTCCAATCGCTGTAAAACTTTCGAAACA
GGCAATCAATAGAGGATTTCAGTGCGACATCGACACGGCTCTGTCGTTTGAGTCCGAA
GCCTTCGGCGAGTGTTTCTCGACGGAAGACCAGAAAGATGCAATGACCGCATTTGTG
GAGAAGAAAAAATCGATGGGTTCAAGAATAGATAA

CsaCRT　ATGGAGTTAAAGAACGTAATCTTGGAAAAGGAGGGCCATCTGGCCATTGTGACGATTA
ATAGACCGAAAGCTTTAAACGCCTTGAACTCGGAGACACTGAAAGACTTGGATACCGT
TATTGAAGACCTTGAAAAGGACTCGAACGTATATAGCGTTATCTTGACTGGCGCAGGC
GAAAAGTCATTCGTGGCAGGAGCAGATATAAGTGAGATGAAAGACTTGAACGAACAG
CAGGGTAAGGAATTTGGGATCTTGGGGAACAATGTCTTTCGGCGTTTGGAAAAGCTTG
ATAAGCCAGTCATCGCTGCCATTAGTGGGTTTGCGTTGGGCGGGGGGTTGTGAATTGG
CAATGAGCTGTGACATTCGTATCGCCTCGGAGAAAGCCAAATTTGGTCAACCTGAGGC
CGGTCTGGGAATAACGCCTGGCTTTGGTGGTACTCAACGCTTAGCGCGTATTGTTGG
GTTAGGCAAGGCAAAAGAGATGATTTATACTTGTGATATAATAAAAGCTGAAGAGGCA
TATCGCATAGGGCTGGTGAATAAGATAGTACCCCTGGAGAACTTAATGGACGAGGCTA
AAGCTATGGCCAATAAAATCATGGCAAATGCACCAAAGGCCGTAAAGTACTGCAAGGA
TGCTATTAATCGGGGTATGCAAGTCGATATTGACGCCGCAATTTTAATTGAAGCTGAG
GATTTCGGTAAATGCTTCGCCACCGAGGATCAAACGGAGGGCATGACCGCATTCCTT
GAAAGAAGAACCGAGAAGAACTTCCAGAACAAGTAA

TdeTER	ATGAAGGTAACCAACCAGAAAGAGCTGAAACAAAAATTAAACGAACTCCGCGAAGCGC
	AAAAAAAATTCGCGACGTATACTCAGGAACAAGTCGATAAGATCTTTAAACAATGTGCC
	ATTGCAGCGGCCAAAGAACGCATCAACCTGGCGAAGTTGGCCGTTGAAGAAACCGGA
	ATTGGTTTAGTGGAAGACAAAATTATTAAGAACCATTTCGCTGCGGAATATATTTATAAT
	AAATACAAAAATGAGAAGACCTGCGGAATTATTGATCATGATGATAGCCTTGGTATCAC
	TAAAGTAGCAGAACCAATCGGTATCGTCGCCGCCATCGTTCCTACAACCAATCCGACC
	TCTACGGCGATCTTTAAATCATTGATTAGCCTGAAAACGCGTAACGCGATTTTTTTCAG
	CCCTCACCCACGCGCCAAAAAAAGCACTATCGCTGCGGCGAAACTGATTCTGGATGC
	GGCAGTTAAAGCCGGCGCACCTAAAAACATTATCGGCTGGATCGACGAGCCTAGCAT
	CGAGTTGAGCCAGGACCTCATGAGTGAAGCAGATATTATCCTCGCCACGGGTGGGCC
	ATCTATGGTTAAAGCGGCCTACTCATCTGGTAAACCAGCCATCGGTGTGGGTGCGGG
	CAATACCCCGGCGATCATTGACGAGAGCGCCGATATTGATATGGCCGTTAGTAGCAT
	CATTCTGAGCAAAACCTACGATAACGGCGTAATTTGCGCGAGTGAACAGAGCATTTTA
	GTGATGAACTCGATCTATGAAAAAGTGAAAGAAGAATTTGTGAAGCGCGGTTCTTACA
	TCCTCAACCAAAATGAAATCGCGAAAATCAAAGAAACGATGTTCAAAAATGGCGCGAT
	CAACGCGGATATTGTTGGCAAATCAGCCTACATTATTGCGAAAATGGCGGGTATTGAA
	GTCCCCCAGACCACAAAGATCCTGATCGGTGAAGTACAGAGCGTCGAAAAGAGCGAG
	CTGTTCAGCCACGAGAAACTGAGCCCTGTTCTGGCCATGTACAAGGTAAAAGATTTTG
	ACGAAGCACTTAAAAAAGCCCAACGCCTTATCGAATTAGGAGGGTCTGGCCACACGA
	GCAGCTTGTACATCGACAGCCAGAATAACAAAGACAAAGTCAAAGAATTCGGCCTTGC
	AATGAAAACTTCTCGCACCTTTATTAATATGCCGTCCAGCCAGGGCGCCTCTGGTGAT
	CTGTACAATTTTGCCATTGCCCCGTCGTTTACCCTGGGGTGTGGGACCTGGGGCGGG
	AATTCGGTATCACAGAACGTCGAACCAAAACACCTGTTGAATATTAAATCCGTGGCAG
	AGCGCCGCGAGAACATGCTGTGGTTCAAAGTCCCTCAGAAAATTTACTTCAAGTACGG
	CTGCCTGCGTTTTGCGCTGAAAGAACTCAAAGACATGAACAAAAACGTGCGTTCATC
	GTTACCGATAAGGACCTGTTTAAACTGGGCTACGTAAACAAAATTACTAAAGTGTTGGA
	CGAAATCGATATTAAATACTCCATTTTTACCGACATTAAGTCAGACCCGACCATCGACA
	GCGTCAAAAAAGGGGCAAAAGAAATGCTGAACTTTGAGCCAGATACGATTATCTCAAT
	CGGTGGGGGCTCGCCTATGGACGCTGCGAAAGTGATGCACCTGCTGTATGAGTACCC
	GGAAGCGGAGATTGAGAACCTGGCCATCAATTTTATGGATATTCGTAAGCGTATTTGC
	AATTTTCCGAAACTTGGGACGAAAGCCATCTCCGTCGCGATTCCGACCACTGCAGGTA
	CGGGCAGCGAAGCCACGCCTTTTGCAGTTATCACGAACGATGAGACCGGTATGAAAT
	ATCCGCTCACCTCGTACGAACTGACCCCAAATATGGCCATCATTGATACCGAACTGAT
	GCTCAATATGCCCCGTAAACTCACCGCAGCCACTGGCATTGACGCACTCGTGCACGC
	CATTGAGGCTTATGTCAGCGTGATGGCGACCGATTACACCGATGAATTAGCTCTCCGT
	GCAATCAAAATGATTTTTAAGTACCTCCCGCGTGCGTACAAAAATGGCACGAATGATAT
	CGAGGCGCGTGAAAAGATGGCTCATGCCAGCAACATCGCCGGTATGGCGTTCGCTAA
	TGCCTTCCTGGGTGTATGCCACAGTATGGCACACAAGCTCGGCGCCATGCATCATGT
	ACCACACGGGATTGCCTGTGCCGTGTTAATCGAGGAAGTCATTAAGTACAATGCCACC
	GATTGCCCGACTAAACAGACCGCCTTTCCGCAGTATAAGAGCCCGAATGCAAAACGTA
	AATACGCCGAGATCGCTGAGTATTTGAATCTCAAAGGAACGAGTGATACTGAGAAAGT
	CACCGCCTTGATCGAAGCCATCAGCAAACTTAAAATCGATCTGTCGATTCCGCAGAAC
	ATTAGCGCGGCCGGTATTAACAAGAAAGATTTTTACAACACGCTGGACAAAATGTCTG
	AACTGGCGTTTGACGATCAGTGCACCACGGCGAACCCGCGCTATCCTCTGATTTCCG
	AGCTCAAGGATATCTACATCAAAAGCTTCTAA
FsuTER	ATGATTATTAAACCACTGATCCGCTCTAATATGTGTATCAACGCGCATCCGAAAGGTTG
	TGCCGCCGACGTGAAACATCAAATCGAGTTCATCAAAAAGAAATTCACGACCCGCTCA
	ATCCCGGCGGACGCGCCAAAAACAGTGTTAGTCCTGGGCTGCTCCACTGGATACGGC
	TTAGCATCACGCATCGTCGCGGCTTTTGGTTACAAGGCTGCAACGATTGGGGTATCGT
	TCGAAAAGAAGGCTCCGACGGAGGAATCGGTGAGAGTCGTGAGAAAACAGGCACC
	CCGGGCTGGTATAACAACATGGCGTTTGATAAGTTCGCGAAGGAAGCCGGTCTGGAT
	GCGGTCACCTTCAACGGTGACGCCTTTAGCCATGAAATGCGTCAGAATGTTATCGATA
	CCCTGAAAAAAATGGGTCGCAAAGTAGATCTCTTGGTCTATTCTGTCGCAAGCTCAGT
	CCGCGTTGATCCAGATAACGGGACCATCTACCGCTCAGTTCTGAAGCCCATCGACAA
	AGTGTTCACCGGGGCGACGATCGATTGCCTGTCTGGTAAGATTTCGACAATTTCGGC

CGAACCTGCGACGGCAGAAGAAGCGGCGAACACGGTCAAAGTGATGGGTGGCGAGG
ATTGGGCGTTGTGGGTGCGCAAACTGAAAGAGGCAGGCGTCCTTGCGGAAGGTGTTA
AAACTGTGGCCTATTCCTATATCGGCCCGAAACTCAGCCACGCTATCTATCGCGACGG
CACTATCGGGGGGTGCCAAAAAACACTTGGAAGCTACGGCTCTTGAACTTAACAAAGAG
CTCCAGAATGATCTCCATGGGGAGGCGTATGTGTCGGTGAATAAAGGTTTAGTGACG
CGCAGCTCAGCAGTGATCCCGATCATTCCGATGTACATTTCGGTTCTGTTTAAAGTCA
TGAAAGAAATGGGCAACCACGAAGGCTGTATTGAACAGATGGAACGCCTGATGACGG
AACGCTTGTATACCGGCTCTAAAGTGCCCACCGACGAAAACCATTTGATCCGTATTGA
CGATTATGAATTGGATCCGAAGGTCCAGGCGGAAGTTGATAAGCGCATGGCTACAGT
GACTCAGGAAAATTTGGCGGAAGTGGGTGATCTGGAAGGATACCGTCACGACTTTTT
GGCAACCAATGGCTTCGATATTGACGGTGTGGACTACGAGGCCGATGTGCAAACGTT
AACCTCAATTTGA

FjoTER	ATGATCATCGAGCCGCGCATGCGCGGTTTTATCTGCCTGACTGCGCATCCGGCGGGA TGTGAACAGAATGTTAAAAATCAGATCGAGTATATTAAATCGAAAGGGGCAATCGCCG GCGCCAAAAAGGTTCTGGTGATCGGCGCATCCACGGGTTTCGGTTTAGCATCCCGTA TCACCAGTGCGTTCGGCTCAGATGCTGCTACGATTGGCGTGTTCTTCGAAAAACCGC CCGTCGAAGGTAAGACAGCGTCGCCAGGGTGGTATAATTCGGCCGCATTTGAGAAAG AGGCACATAAAGCGGGTCTTTACGCTAAATCTATCAATGGAGACGCTTTCAGCAACGA AATTAAACGTGAAACCTTAGATCTGATCAAAGCGGATTTAGGTCAGGTTGATCTGGTAA TTTATTCGCTGGCGTCCCCGGTTCGTACGAACCCGAACACAGGTGTGACTCACCGCA GTGTGTTGAAACCGATCGGTCAGACTTTTACAAACAAACTGTGGATTTTCATACGGG GAACGTGTCCGAAGTTTCTATCGCGCCGGCTAATGAAGAAGATATTGAAAATACGGTA GCAGTGATGGGCGGAGAAGATTGGGCGATGTGGATTGATGCCCTCAAAAATGAAAAT CTGCTGGCAGAGGGGGCGACGACAATTGCATATTCCTATATTGGCCCGGAATTGACC GAAGCGGTCTACCGTAAAGGCACCATTGGTCGTGCAAAAGACCACCTGGAGGCGACC GCTTTCACCATTACTGATACCCTTAAATCGTTAGGCGGAAAAGCGTACGTGTCGGTGA ATAAAGCCTTGGTTACGCAAGCCTCGTCGGCGATTCCTGTGATCCCGCTGTATATCTC GCTGCTTTATAAAATTATGAAGGAGGAAGGAATTCACGAGGGATGCATCGAACAAATT CAGCGCTTGTTCCAAGATCGTTTGTATAACGGTAGCGAAGTGCCGGTTGATGAGAAAG GCCGCATCCGCATTGACGATTGGGAGATGCGCGAGGATGTGCAGGCTAAAGTTGCG GCTCTGTGGAAGGAAGCCACCACCGAAACCCTGCCATCCATCGGCGACCTGGCAGG TTACCGTAATGACTTCTTAAACCTGTTTGGGTTTGAATTTGCGGGAGTGGATTACAAGG CGGATACGAACGAGGTCGTAAACATTGAAAGCATCAAATAA
CacADH	ATGAAGGTAACCAACCAGAAAGAGCTGAAACAAAAATTAAACGAACTCCGCGAAGCGC AAAAAAAAATTCGCGACGTATACTCAGGAACAAGTCGATAAGATCTTTAAACAATGTGCC ATTGCAGCGGCCAAAGAACGCATCAACCTGGCGAAGTTGGCCGTTGAAGAAACCGGA ATTGGTTTAGTGGAAGACAAAATTATTAAGAACCATTTCGCTGCGGAATATATTTATAAT AAATACAAAAATGAGAAGACCTGCGGAATTATTGATCATGATGATAGCCTTGGTATCAC TAAAGTAGCAGAACCAATCGGTATCGTCGCCGCCATCGTTCCTACAACCAATCCGACC TCTACGGCGATCTTTAAATCATTGATTAGCCTGAAAACGCGTAACGCGATTTTTTTCAG CCCTCACCCACGCGCCAAAAAAAGCACTATCGCTGCGGCGAAACTGATTCTGGATGC GGCAGTTAAAGCCGGCGCACCTAAAAACATTATCGGCTGGATCGACGAGCCTAGCAT CGAGTTGAGCCAGGACCTCATGAGTGAAGCAGATATTATCCTCGCCACGGGTGGGCC ATCTATGGTTAAAGCGGCCTACTCATCTGGTAAACCAGCCATCGGTGTGGGTGCGGG CAATACCCCGGCGATCATTGACGAGAGCGCCGATATTGATATGGCCGTTAGTAGCAT CATTCTGAGCAAAACCTACGATAACGGCGTAATTTGCGCGAGTGAACAGAGCATTTTA GTGATGAACTCGATCTATGAAAAAGTGAAAGAAGAATTTGTGAAGCGCGGTTCTTACA TCCTCAACCAAAATGAAATCGCGAAAATCAAAGAAACGATGTTCAAAAATGGCGCGAT CAACGCGGATATTGTTGGCAAATCAGCCTACATTATTGCGAAAATGGCGGGTATTGAA GTCCCCCAGACCACAAAGATCCTGATCGGTGAAGTACAGAGCGTCGAAAAGAGCGAG CTGTTCAGCCACGAGAAACTGAGCCCTGTTCTGGCCATGTACAAGGTAAAAGATTTTG ACGAAGCACTTAAAAAAGCCCAACGCCTTATCGAATTAGGAGGGTCTGGCCACACGA GCAGCTTGTACATCGACAGCCAGAATAACAAAGACAAAGTCAAAGAATTCGGCCTTGC AATGAAAACTTCTCGCACCTTTATTAATATGCCGTCCAGCCAGGGCGCCTCTGGTGAT CTGTACAATTTTGCCATTGCCCCGTCGTTTACCCTGGGGTGTGGGACCTGGGGCGGG

AATTCGGTATCACAGAACGTCGAACCAAAACACCTGTTGAATATTAAATCCGTGGCAG
AGCGCCGCGAGAACATGCTGTGGTTCAAAGTCCCTCAGAAAATTTACTTCAAGTACGG
CTGCCTGCGTTTTGCGCTGAAAGAACTCAAAGACATGAACAAAAACGTGCGTTCATC
GTTACCGATAAGGACCTGTTTAAACTGGGCTACGTAAACAAAATTACTAAAGTGTTGGA
CGAAATCGATATTAAATACTCCATTTTTACCGACATTAAGTCAGACCCGACCATCGACA
GCGTCAAAAAAGGGGCAAAAGAAATGCTGAACTTTGAGCCAGATACGATTATCTCAAT
CGGTGGGGGCTCGCCTATGGACGCTGCGAAAGTGATGCACCTGCTGTATGAGTACCC
GGAAGCGGAGATTGAGAACCTGGCCATCAATTTTATGGATATTCGTAAGCGTATTTGC
AATTTTCCGAAACTTGGGACGAAAGCCATCTCCGTCGCGATTCCGACCACTGCAGGTA
CGGGCAGCGAAGCCACGCCTTTTGCAGTTATCACGAACGATGAGACCGGTATGAAAT
ATCCGCTCACCTCGTACGAACTGACCCCAAATATGGCCATCATTGATACCGAACTGAT
GCTCAATATGCCCCGTAAACTCACCGCAGCCACTGGCATTGACGCACTCGTGCACGC
CATTGAGGCTTATGTCAGCGTGATGGCGACCGATTACACCGATGAATTAGCTCTCCGT
GCAATCAAAATGATTTTTAAGTACCTCCCGCGTGCGTACAAAAATGGCACGAATGATAT
CGAGGCGCGTGAAAAGATGGCTCATGCCAGCAACATCGCCGGTATGGCGTTCGCTAA
TGCCTTCCTGGGTGTATGCCACAGTATGGCACACAAGCTCGGCGCCATGCATCATGT
ACCACACGGGATTGCCTGTGCCGTGTTAATCGAGGAAGTCATTAAGTACAATGCCACC
GATTGCCCGACTAAACAGACCGCCTTTCCGCAGTATAAGAGCCCGAATGCAAAACGTA
AATACGCCGAGATCGCTGAGTATTTGAATCTCAAAGGAACGAGTGATACTGAGAAAGT
CACCGCCTTGATCGAAGCCATCAGCAAACTTAAAATCGATCTGTCGATTCCGCAGAAC
ATTAGCGCGGCCGGTATTAACAAGAAAGATTTTTACAACACGCTGGACAAAATGTCTG
AACTGGCGTTTGACGATCAGTGCACCACGGCGAACCCGCGCTATCCTCTGATTTCCG
AGCTCAAGGATATCTACATCAAAAGCTTCTAA