Engineering Biosynthetic Pathways in Cell-Free Systems for Sustainability and Chemical Innovation

Thesis by

Yong Yi Wu

In Partial Fulfillment of the Requirements for the

Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2018

(Defended August 7, 2017)

Yong Yi Wu

ORCID: 0000-0002-5401-3662

In Memory

This manuscript is dedicated to

Mrs. Nancy Carter Svendsen

(1943-2007)

a dedicated and benevolent teacher

She has shown me a path towards new light.

ACKNOWLEDGEMENTS

I would like to thank Professor Mikhail G. Shapiro and Professor David A. Tirrell for serving as my committee members. Their thoughtful insights and guiding questions have helped shape my research projects profoundly.

I would like to thank my advisor Professor Richard M. Murray for having me to join his group. He matches my skills and interests with my research projects, cultivates the development of my leadership skills, and fosters my professional growth. He had stood by me when I was weak. The serendipity of meeting him has made my goal for graduate school a reality: integrating ideas and innovations from multiple disciplines and contributing to the advancement of biotechnology. It has been a great pleasure to work with many talented individuals in the Murray Lab. I especially want to thank Jongmin Kim, Yutaka Hori, Joseph Meyerowitz, Emzo de los Santos, Enoch Yeung, Zachary Sun, Claremyra Hayes, Vipul Singhal, Anu Thubagere, Victoria Hsiao, Zoltan Tuza, and Shaobin Guo who have shaped me into the researcher I am today. I want to thank Scott C. Livingston, Anandh Swaminathan, and James Parkin for being supportive officemates.

I would like to thank Professor Frances H. Arnold for collaborating with me and pushing me to think critically. She has served as the role model who continually inspires me to strive for excellence. She has guided me through hard times and helped me get back on my feet. Her sharp intuitions and kind encouragements have been extremely motive and valuable to my Ph.D. experience. The people in the Arnold group have been very constructive in directing my ideas by teaching me directed evolution and organic chemistry. I would especially like

to thank individuals who have directly interacted with me: Sheel Dodani, David Romney, Andrew Buller, Russell Lewis, and Anders Knight.

I want to thank the people at Genomatica who collaborated with me. All of them have created a very welcoming and educational environment for me to succeed as a visiting scientist and a Ph.D. student. They have shown me the wonder of team building and knowledge transfer. Special thanks to those who have interacted with me directly: Hirokazu Sato, Eric Ralph, Jingyi Li, Sari Rizek, Glenn Majer, Hongjun Huang, Evan Ehrich, Emily Mitchell, Anna Lechner, Joseph R. Warner, Jonathan Joaquin, Jungik Choi, John Trawick, Harry Yim, Jeff Bolt, Tae Hoon Yang, Julia Khandurina, Robert Haselbeck, Robin Osterhout, Stephanie Culler, Cara Tracewell, and Stephen Van Dien. It was a great pleasure to interact with Hiro, Stephanie, and Steve on a regular basis. Their mentorship, support, and positivity has guided me through my project and helped me grow professionally. Hiro's strong microbiology background, Steve's deep scientific knowledge, and Stephanie's sharp visionary mind have helped me excel as a researcher tremendously. The rare opportunity to work alongside Stephanie and Steve has allowed me to experience and appreciate effective management styles and admirable leadership qualities, which has benefited me greatly.

I want to thank Dr. Nathan F. Dalleska and the Environment Analysis Center. He has taught me different ways to analyze compounds using various instruments named with one acronym after another. In five years, he has helped me detect more than a hundred compounds on five different machines. His dedication, patience, and wit have guided me through many bumps and roadblocks. His occasional life advice has also enlightened me in many ways. I also want

to thank the Grubbs Lab and especially Zachary Wickens, who let me use their GC-FID for my initial project at the Murray Lab. Thanks to Professor Robert H. Grubbs' generous trust in me, I was able to establish the basis of my Ph.D. work with the help of his lab's expertise and resources.

I thank those who have always supported me. I thank the administrators that have helped me along the way: Kathy Bubash, Liz Ayala, and Nikki Fountleroy. Liz, along with Professor Paul W. Sternberg, has been my rock and helped me survive the loss of my laptop and granted me the financial opportunity to set up high-throughput screening for my projects. I thank those who lend a helping hand and those who lend a pair of patient ears. Joy Doong and Nancy Cao were the best research assistances I could have ever asked for in times of need. I also want to thank David Seung and his families for supporting my travel between Lo Jolla and Pasadena. My mentors along the way have made a difference during my time at Caltech, including Soyoung Park, Peter Rapp, and Zuleikha Kurji. Camille McAvoy, Lucy Ji, and Karthik Ramachandran have been real friends since the first day I came to Pasadena, CA. I would not have made it through graduate school without them. Thanks to them, I managed to overcome one challenge after another and forgive past shortcomings. I also want to thank my classmates who have supported me through the chemical engineering curriculum.

I want to thank all the teaching assistants, mentors, counselors, teachers, lecturers, and professors from graduate school, college, high schools, middle school, and elementary school that have guided my journey towards a Ph.D. Special thanks to Ms. Kelly, who showed me a path to pursue higher education. Special thanks to Luann M. Pugh, Clarence G. Hermansky,

and Jingjing Zhang, who have helped me make my decision to apply to graduate schools in the first place.

I am forever indebted to my sister Ying Wu and my mother Mei Liang who have supported my decision to pursue higher education. Many thanks to Mrs. Nancy C. Svendsen (1943-2007), who has believed in me from the very beginning of my quest for a better life in the United States; I would not have made it this far without her kind advice. Thanks to UVA President John T. Casteen III, who created the AccessUVA program, and Carolyn Vallas, who directed the UVA Center for Diversity in Engineering, both of which have fully supported my pursuit of a bachelor degree in chemical engineering. I would not have made it this far without the generous supports from the following organizations (in chronological orders). I thank those that have supported me through undergraduate studies: Horatio Alger Association, Jefferson Prep, Asian & Pacific Islander American Scholarship Fund, Lisa Sechrist Memorial Foundation, Apple Federal Credit Union, Ronald McDonald House Charities, Stuart Foundation, AccessUVa, and SEAS Class Of 1986. I also thank those that have supported my graduate studies: The Department of Chemical Engineering at Caltech, DARPA Living Foundries Program, NIH/NRSA Training Grant, Genomatica, and Gordon and Betty Moore Foundation.

ABSTRACT

This work presents the cell-free transcription-translation (TX-TL) system as a research and development platform for renewable synthesis and molecular discovery. TX-TL is easy to use and provides a biomolecular breadboard for the rapid prototyping and engineering of biosynthetic pathways. This work has validated the capabilities of the cell-free TX-TL system for simultaneous protein expression and chemical synthesis. Specifically, this work shows that TX-TL supports the conversion of intermediates from carbohydrate metabolism and amino acids into valuable compounds. Metabolic flux through cofactor dependent pathways confirms that active cofactor metabolism is occurring in TX-TL. This work has also demonstrated the industrial relevance of TX-TL through exploring design space of a biosynthetic pathway for improved product yield and expanding substrate scope of another biosynthetic pathway.

Current methods for assembling biosynthetic pathways in microorganisms require a process of repeated trial and error and have long design-build-test cycles. We describe the use of a cell-free transcription-translation (TX-TL) system as a biomolecular breadboard for the rapid engineering of the 1,4-butanediol (BDO) pathway. In this work, we have verified enzyme expression and enzyme activity and identified the conversion of 4-hydroxybutyrate to downstream metabolites as the pathway bottleneck. We demonstrate the reliability of using linear DNA in TX-TL as a tool for engineering biological systems by undertaking a careful characterization of its transcription and translation capabilities and provide a detailed analysis of its metabolic output. Pathway constructs of varying pathway

enzyme expression levels are tested in TX-TL and *in vivo* to identify correlations between the two systems, and we find that the production of BDO is correlated to the expression of enzyme *ald* in both systems. The use of TX-TL to survey the design space of the BDO pathway enables rapid tuning of pathway enzyme expression levels for improved product yield. Different pathway combinations are also tested in TX-TL for its application in pathway ranking. Leveraging TX-TL to screen enzyme variants for improved catalytic activity accelerates design iterations that can be directly applied to *in vivo* strain development.

TX-TL simulates a customizable cellular environment that can be controlled by manipulating pH, changing cellular components, or adding exogenous substrates. By adding linear DNA encoding individual enzymes of the violacein pathway and tryptophan analogs in TX-TL reactions, we have discovered new violacein analogs. TX-TL enables rapid production of natural product analogs with diverse substitution, which allows small-scale biosynthesis of potential drug candidates and offers a new platform for drug discovery. This work also presents TX-TL as a platform for protein engineering. Residues targeted for site-saturated mutagenesis were identified with protein-ligand docking. Linear DNAs of individual enzyme mutants were added into TX-TL reactions to screen for improved enzyme variant. Screening result indicates vioE mutant Y17H reduces byproduct formation and redirects metabolic flux towards target metabolites. Protein engineering for improved enzyme activity can further expand the substrate scope of a natural product pathway and result with more natural product analogs that can be applied for medical applications.

This work demonstrates that the cell-free TX-TL system can become a valuable tool that complements the process of engineering biosynthesis in the whole cell *in vivo* system or the purified protein *in vitro* system. Future engineering and development of the TX-TL system can further expand the chemical space for biosynthesis.

PUBLISHED CONTENT AND CONTRIBUTIONS

Nguyen, P. H. B.; Wu, Y.; Guo, S.; Murray, R. M., Design Space Exploration of the Violacein Pathway in Escherichia coli Based Transcription Translation Cell-Free System (TX-TL). bioRxiv **2015**. DOI: 10.1101/027656

Y. Y. W. initiated, defined, and designed the project. Y.Y.W. designed primers necessary for cloning, sequencing, and linearizing DNA, developed experimental procedures for extracting products for analysis, performed Western blotting, operated the LC-MS, developed calibration curve, quantified products, debugged MATLAB code, maintained lab supplies, participated in data analysis, and edited the final manuscript.

Wu, Y. Y.; Culler, S.; Khandurina, J.; Van Dien, S.; Murray, R. M., Prototyping 1,4-butanediol (BDO) biosynthesis pathway in a cell-free transcription-translation (TX-TL) system. bioRxiv **2015**. DOI: 10.1101/017814

Y. Y. W participated in the conception of the project, designed primers and experimental procedures, executed all experiments, analyzed results, and wrote the manuscript.

Wu, Y. Y.; Sato, H.; Huang, H.; Culler, S.; Khandurina, J.; Nagarajan, H.; Yang, T. H.; Van Dien, S.; Murray, R. M., System-level studies of a cell-free transcription-translation platform for metabolic engineering. bioRxiv **2015**. DOI: 10.1101/172007

Y. Y. W participated in the conception of the project, designed primers and experimental procedures, executed all TX-TL experiments, analyzed results, and wrote the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

NOMENCLATURE

Acetyl-coA. Acetyl coenzyme A.

ACT. Acetoin.

ATP. Adenosine triphosphate.

coA. Coenzyme A.

4-HB. 4-hydroxybutyrate.

BCD. Bicistronic design.

BDO. Butanediol. 1,4-BDO implies 1,4-butanediol, and 2,3-BDO implies 2,3-butanediol.

cAMP. Adenosine-3′,5′-cyclic monophosphate.

CPA. Chromopyrrolic acid.

CTP. Cytidine triphosphate.

CV. Coefficient of variation.

DNA. Deoxyribonucleic acid.

DTT. 1,4-Dithiothreitol.

ESI. Electrospray ionization.

FBA. Flux Balance Analysis.

GBL. γ-Butyrolactone.

GC. Gas Chromatography.

GFP. Green fluorescent protein.

HEPES. 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid.

HRMS. High resolution mass spectrometry.

IPA. Indole-3-pyruvate.

IPTG. Isopropyl-β-D-thiogalactopyranoside.

KOH. Potassium hydroxide.

LC. Liquid Chromatography.

Linear DNA. Deoxyribonucleic acid in linear forms.

MALDI. Matrix-assisted laser desorption/ionization.

MOPS. 4-Morpholinepropanesulfonic acid.

MRM. Multiple reaction monitoring.

mRNA. Messenger RNA.

MS. Mass Spectrometry.

NAD. Nicotinamide adenine dinucleotide.

NADH. The reduced form of NAD and the reduced cofactor for the synthesis of 1,4-BDO and 2,3-BDO.

NADP. Nicotinamide adenine dinucleotide phosphate.

NADPH. The reduced form of NADP and the reduced cofactor for the synthesis of 1,4-BDO and violacein.

PCR. Polymerase chain reaction.

PDB. Protein Data Bank.

PEG. Polyethylene glycol.

RBS. Ribosomal binding site.

RNA. Ribonucleic acid.

RT-PCR. Reverse transcription polymerase chain reaction.

TCA cycle. Tricarboxylic acid cycle.

tRNA. Transfer RNA.

TX-TL. Transcription-Translation, which refers to the cell-free system used in this work.

UTP. Uridine triphosphate.

UTR. Untranslated region.

Chapter 1

INTRODUCTION

METABOLIC ENGINEERING

Advancements in the engineering of biological systems for valuable compounds have formed the cornerstone of the biotechnology industry. Bio-based processes utilizes renewable resources to produce sustainable chemicals, which have become a very attractive alternative to traditional organic synthesis.[1-2] The bio-based production of C2 to C6 platform chemicals[3] provides evidence that the field of bioengineering can develop tools for building the future of green chemistry. The research and development efforts towards producing economically competitive chemicals in biological systems have resulted in a wide range of techniques to maximize target metabolite production. Two main approaches for the engineering of biosynthesis are: to increase fluxes through target pathway reactions and to improve target pathway enzymes' performance.[4] Some representative techniques for the first approach include flux balance analysis (FBA),[5] and ^{13}C isotope tracing for flux determination.[6] The flux-emphasized approach optimizes target compound production by identifying the most or the least important reactions of a particular biosynthetic pathway. The most critical reactions can be up-regulated by overexpressing related enzymes, while the least desirable reactions can be eliminated by gene knockout. For example, the engineering of *E. coli* for the production of 1,3-propanediol from glucose was achieved by removing nonproductive genes and the energetically inefficient glucose transport mechanism.[7]

Instead of rerouting the metabolic network, the second approach aims to increase target metabolite production by protein engineering. Protein engineering can be achieved by two main strategies: rational design and directed evolution.[8] By introducing genetic mutation to a targeted enzyme, enzyme variants can facilitate novel chemistry. For example, a cytochrome engineered with unique serine-heme ligation catalyzes olefin cyclopropanation in *E. coli*.[9] Other techniques for the engineering of biosynthesis includes enzyme co-localization[10-12] and the implementation of metabolite sensors and regulators.[13-15] Many have also started to expand the chemical space by looking at biosynthesis with new lenses. Some of the ongoing efforts include studies on human gut biota,[16] pathway mining,[17] and the construction of *de novo* pathway.[18] Collectively, the field of metabolic engineering has the great potential to transform bio-based production to compete with the conventional chemical processes. In this work, techniques such as flux analysis, expanding substrate scope, and protein engineering are applied to explore the capabilities of cell-free systems as a platform for the engineering of biosynthesis.

PROTOTYPING IN A CELL-FREE BIOMOLECULAR BREADBOARD

Cell-free systems offer a reliable testing environment for engineering biosynthetic pathways. Cell membranes are essential for protecting cells from the surrounding environment and regulating material exchange, but cell membranes sometimes impose constraints for engineering biological systems.[19] Removing cell membranes from organisms and retaining cytoplasmic contents creates cell-free systems. The process of making cell extract breaks down cell membrane, gets rid of endogenous DNA, and retains cellular components

necessary for transcription and translation. Building blocks for transcription and translation (nucleoside triphosphate and amino acids) and energy currency (cofactors such as coA and NAD) are added into the extract to facilitate biochemical reactions. The homogeneity of the cell extracts enables better control of the reaction environment and easy access to cellular components.[20] Cell-free biology opens up opportunities to gain biological insights and potentially create technological innovations. Cell-free transcription and translation systems were developed in the early 1980s.[21-23] Since then, researchers have been utilizing the systems for various applications. Cell-free protein expression coupled with large-scale bioprocessing can introduce new ways for pharmaceutical productions. For example, Sutro BioPharma is currently in the process of scaling up cell-free production of pharmaceutical proteins.[24] The applications of cell-free systems for biological engineering has renewed interests.[25] Cell-free transcription-translation system (TX-TL) has been applied as a biomolecular breadboard for efficient biological circuit prototyping.[26-29] There are also cell-free systems engineered specifically for the continuous synthesis of renewable energy and materials.[30-32]

Cell-free systems simulate a controlled cellular environment and are suitable for prototyping biological parts and modules. Cell-free systems decouple growth from the active transcription-translation machinery, which eliminates cell transformation, cell culture seed train, and the application of antibiotics. Other advantages of using cell-free systems as a biomolecular breadboard include open access to the cellar biochemical network, protein expressions from linear pieces of DNAs, and simultaneous protein expressions of multiple pieces of DNAs. Open access to the cellular network allows the addition of exogenous

materials and subtraction of undesired cellular components, which creates opportunities to explore non-natural enzyme activity that may be otherwise impossible *in vivo*. Protein expressions from multiple linear pieces enable high-throughput screening of different pathway combinations, pathway compartments, and cofactor specificities. This feature significantly reduces cloning efforts necessary for prototyping. Collectively, cell-free systems accelerate the design-build-test cycle for engineering biosynthesis pathways. The cycle time for a biological part assembly can be significantly reduced from 1 week (Figure 1.1) *in vivo* to 1 day using linear DNAs in the cell-free TX-TL system.[33] Further, the dynamics of a biosynthetic pathway can be analyzed in less than eight hours, which is a fraction of the time it takes for total biosynthesis *in vivo* (shown in Figure 1.2). In this work, Chapter 2 shows the benefits of the cell-free TX-TL system and presents a systematic comparison between TX-TL and *in vivo* system. Appendix A also showcases a piece of earlier work on prototyping biosynthetic pathways in TX-TL.

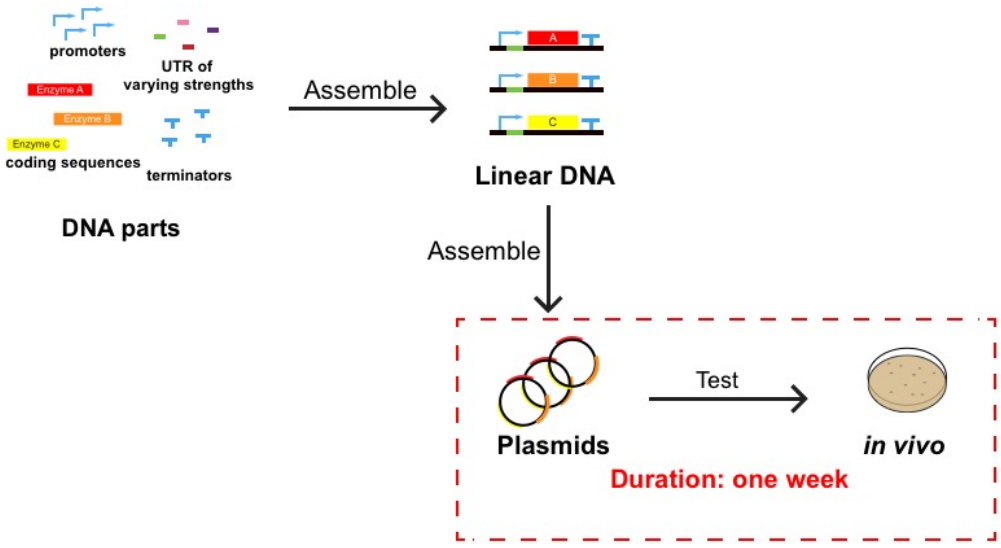


Figure 1.1: Design-build-test cycle for biosynthesis. The testing duration from plasmid to metabolite *in vivo* takes up to one week.

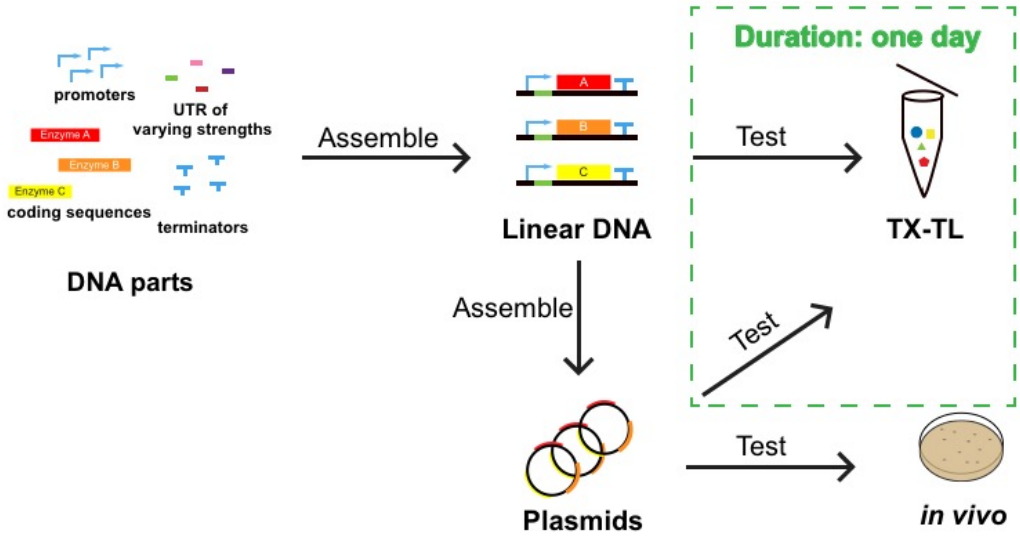


Figure 1.2: New design-build-test cycle for biosynthesis. The testing duration from plasmid or even linear DNA to metabolite takes one day.

RESOURCE LIMITATIONS and DESIGN SPACE EXPLORATION

Cellular resource limitation poses challenges to the engineering of biosynthesis.[34] Previous research has studied the central metabolism and rerouted metabolic pathways for improved pathway yields.[35-36] Nevertheless, protein synthesis is an energy-intensive cellular process. The allocation of resources between metabolite synthesis and protein synthesis remains an optimization problem for biosynthesis.[37] Further, protein expression levels and activity affect target product yield. Tuning protein expression levels requires engineering on multiple levels, including transcription from DNA to RNA, translation from RNA to protein, and protein activity. A biosynthetic pathway facilitates a series of reactions from raw materials to target products with the help of multiple proteins, and balancing multiple protein expressions in parallel remains a challenge. Conventionally, enzymes are overexpressed when they are identified to be important for improving target metabolite production.

However, protein overexpression can affect cell growth[38-39] and possibly reduce metabolite production. It will be advantageous to predict and test for pathway enzymes' protein expression levels that result with optimal metabolite production.

Metabolic pathways consisting of multiple parts and factors should be combined in precise combinations to achieve desired functions. Simultaneous protein expression of multiple pieces of DNA in cell-free system facilitates design space exploration of metabolic pathways. In the cell-free TX-TL system, the ratio of pathway enzyme expressions can be adjusted by varying the concentration of plasmid encoding individual pathway enzymes. Subsequently, the ratios of pathway enzyme expressions can be adjusted for balancing metabolic flux. In this work, Chapter 2 investigates the resource limitation of TX-TL and its impact on a metabolic pathway. Strategies to address resource limitation and balance gene expression for the biosynthesis of 1,4-butanediol in the cell-free TX-TL system is presented, as well as a systematic comparison of TX-TL to small-scale *in vivo* systems.

PROTEIN ENGINEERING AS A TOOL FOR DRUG DISCOVERY

Driven by chemistry, drug discovery contributed to the progress of medicine significantly in the past century.[40] Chemical diversity for drug screening[41] can be achieved by biosynthesis, and synthetic biology is reorienting the field of drug discovery with the development of new tools.[42] These recent technological advances have renewed interests in natural products.[43] Natural products are a set of diverse compounds with great potential for medical applications.[44] From 1940s to 2010, natural products and their analogs make up 48.6% of anti-cancer small molecules.[45] Recent advancements in the biosynthesis of natural products

include the synthesis of anti-malaria drug artemisinin.[46] The compound was first isolated from traditional Chinese herb medicine for malaria treatments.[47] Some natural products can be directly used for effective medical applications, others require modifications to their chemical structures. The addition of functional groups such as methyl, fluoro, and nitrile to natural products can improve compound rigidity,[48] modulate cation-π interaction,[49] and polarize adjacent electron density,[50] respectively. Such modifications can enhance the potency, lipophilicity, bioavailability, and metabolic stability of a drug candidate.[51-52]

The development of protein engineering using directed evolution has led to significant breakthrough in the biosynthesis of drug candidates.[53] Although organic synthesis for natural products and their analogs remains a challenge,[54] naturally-occurring enzymes can be engineered to outcompete organometallic catalysts. For example, P450 was engineered for enantioselective formal synthesis of levomilnacipran.[55] Further, the engineering of an enzyme for non-natural activities can usually open up possibilities for the synthesis of a new array of compounds. For example, the engineering of TrpB leads to the production of an array of non-natural tryptophan with the additional functional groups.[56] In this work, Chapter 3 presents the cell-free TX-TL system as a platform for protein engineering and substrate scope expansion for drug discovery.

Chapter 2

SYSTEM-LEVEL STUDIES OF A CELL-FREE TRANSCRIPTION-TRANSLATION PLATFORM FOR METABOLIC ENGINEERING

Yong Y. Wu[*], Hirokazu Sato[‡], Hongjun Huang[‡], Stephanie J. Culler[‡], Julia Khandurina[‡], Harish Nagarajan[‡], Tae Hoon Yang[‡], Stephen Van Dien[‡], and Richard M. Murray[*]

[*]: California Institute of Technology, 1200 E. California Blvd., Pasadena CA 91125

[‡]: Genomatica Inc., 4757 Nexus Center Dr., San Diego, CA 92121

Keywords: cell-free systems, metabolic engineering, design space exploration, and 1,4-butanediol

ABSTRACT

Current methods for assembling biosynthetic pathways in microorganisms require a process of repeated trial and error and have long design-build-test cycles. We describe the use of a cell-free transcription-translation (TX-TL) system as a biomolecular breadboard for the rapid engineering of the 1,4-butanediol (BDO) pathway. In this work, we have verified enzyme expression and enzyme activity and identified the conversion of 4-hydroxybutyrate to downstream metabolites as the pathway bottleneck. We demonstrate the reliability of using linear DNA in TX-TL as a tool for engineering biological systems by undertaking a careful characterization of its transcription and translation capabilities and provide a detailed analysis of its metabolic output. Pathway constructs of varying pathway enzyme expression levels are tested in TX-TL and *in vivo* to identify correlations between the two systems, and we find that the production of BDO is correlated to the expression of enzyme *ald* in both systems. The use of TX-TL to survey the design space of the BDO pathway enables rapid tuning of pathway enzyme expression levels for improved product yield. Different pathway combinations are also tested in TX-TL for its application in pathway ranking. Leveraging TX-TL to screen enzyme variants for improved catalytic activity accelerates design iterations that can be directly applied to *in vivo* strain development.

INTRODUCTION

Utilizing fast-growing microorganisms to produce molecules of industrial relevance has the potential to advance the progress of green chemistry rapidly. Processes of traditional chemical synthesis require heavy metal catalysts, toxic solvents, and fossil fuels as feedstocks. The biosynthetic approach, which uses naturally occurring enzymes, less energy, and renewable feedstocks, is becoming an attractive alternative.[1, 3] However, biosynthetic approaches are challenged by long design-build-test cycles. Microbial pathway engineering often has about one-week cycle time.[57] The performance of the pathway is frequently far from design, requiring many iterations.[58] For example, it took DuPont and Genencor more than 100 person-years of work to develop the commercialization of bio-based 1,3-propanediol.[59] Recent advances in cell-free systems offer an alternative to this costly approach. Cell-free systems have been used to reduce the cycle time of pathway construction. The design-build-test cycle in a cell-free system using linear DNA takes less than one day.[60]

Cell-free systems simulate a controlled cellular environment that delivers repeatable results. Recent research has explored the application of cell-free systems for biological circuits, renewable energy, and medicine. The cell-free transcription-translation (TX-TL) system was first developed as a biomolecular breadboard to test genetic circuits, and many have been demonstrated since.[26-29] The synthesis of hydrogen and the development of enzymatic fuel cells in cell-free systems has charted new paths for renewable energy.[30-31] The high yield of therapeutic proteins in *E. coli*-based cell-free synthesis system also offered new methods for medicinal synthesis.[24, 61-62] Using cell-free systems for prototyping metabolic pathways is an

attractive alternative platform for the engineering of biosynthesis in microbial hosts. Lysate of engineered *E. coli* has been used to support high-level conversion of valuable chemicals,[63] and a cell-free protein synthesis system has been used to screen enzyme variants.[64] Systematic studies of cell-free systems have monitored the change of pH, measured the change of metabolites over time, and evaluated protein synthesis.[20, 65-66] Studies have shown that the depletion of ATP is limiting near the beginning of a cell-free reaction, and the consumption of glutamate is critical in regenerating cofactors.[67-69] A correlation between cell-free and *in vivo* systems has not been demonstrated. As such, a systematic side-by-side analysis between *in vivo* and cell-free systems is required for cell-free systems to be considered widely as a prototyping platform for metabolic engineering.

Metabolic pathways consist of multiple parts and factors that are combined in precise combinations to achieve desired functions. Enzyme expression levels and activity affect target product yield. Tuning enzyme expression requires engineering the level of transcription, translation, and enzyme activity. Furthermore, balancing expression of multiple genes in parallel remains a challenge. Often, enzymes are overexpressed when they are identified to be essential for improving target metabolite production. Protein overexpression can affect cell growth[39] and possibly reduce metabolite production.[70] Studies have previously demonstrated the feasibility of tuning protein expression levels *in vivo* for improving metabolite productions.[71-73] Cell-free TX-TL system also provides a platform for investigating the correlation between protein expression levels and metabolite production. Cell-free TX-TL system allows simultaneous protein expression from multiple pieces of DNA, including linear DNA. Such properties can be used to verify pathway enzyme

expression and activity. Sun *et al.*'s work has connected protein expression level in TX-TL to *in vivo* systems by comparing the expression strength of different promoters.[60] This work takes a further step to compare the dynamics of a metabolic pathway in TX-TL to *in vivo* systems by correlating protein expression levels to metabolite production in both systems.

This work aims to demonstrate the reduction of traditional metabolic engineering design-build-test cycles using the 1,4-butanediol (1,4-BDO) pathway as a TX-TL prototype. 1,4-BDO and its derivatives are widely used for producing automotive plastics, electronic chemicals, and elastic fibers. 1,4-BDO has a projected global market of $8.96 billion by 2019.[74] Historically, 1,4-BDO has been produced from petrochemical feedstocks, but recently a bio-based process was commercialized.[75] The 1,4-BDO pathway used for this bioprocess is shown in Figure 2.1. The pathway converts a tricarboxylic acid (TCA) cycle intermediate succinyl-coA to 1,4-BDO. From top to bottom of the pathway schematic, pathway intermediates include succinyl semialdehyde, 4-hydroxybutyrate (4HB), 4-hydroxybutyryl-coA (4HB-coA), and 4-hydroxybutryaldehyde (4HB-aldehyde). CoA-dependent succinate semialdehyde dehydrogenase (*sucD*) catalyzes the conversion of succinyl-coA to succinyl semialdehyde. 4- hydroxybutyrate dehydrogenase (*4-hbd*) catalyzes succinyl-coA to 4HB. 4-hydroxybutyryl-coA transferase (*cat2*) catalyzes 4HB to 4HB-coA. 4-hydroxybutyryl-coA reductase (*ald*) catalyzes 4HB-coA to 4HB-aldehyde. Alcohol dehydrogenase (*adh*) catalyzes 4HB-aldehyde to 1,4-BDO.

The main goal of this work is to demonstrate TX-TL as a research tool for metabolic engineering and to establish the feasibility of design space exploration (shown in Figure 2.1).

Using the 1,4-BDO pathway previously developed by Genomatica Inc.,[76] we add linear DNA encoding pathway enzymes to TX-TL. We measure the resulting transcriptional and translational outputs and pathway related metabolites. We also use TX-TL to rapidly tune pathway enzyme expression levels for design space exploration of the 1,4-BDO pathway. Ribosome-binding site (RBS) elements of varying strengths are chosen from the bicistronic design (BCD) library to adequately explore the design space of the 1,4-BDO pathway in TX-TL. The two Shine-Dalgarno motifs from BCD allows the first one to make a leader peptide to open the second one, which delivers precise and reliable translation initiation. The translational coupling architecture BCD ensures protein expression at expected levels independent of downstream sequence.[77] Through exploring the design space of the 1,4-BDO pathway in TX-TL and *in vivo*, we systematically compare the metabolic output and enzyme expression levels. To show the industrial relevance of TX-TL, we demonstrate that the use of linear DNA in TX-TL has the capabilities to serve as a biomolecular breadboard to speed up design iterations, and results from TX-TL can be applied to *in vivo* strain development.

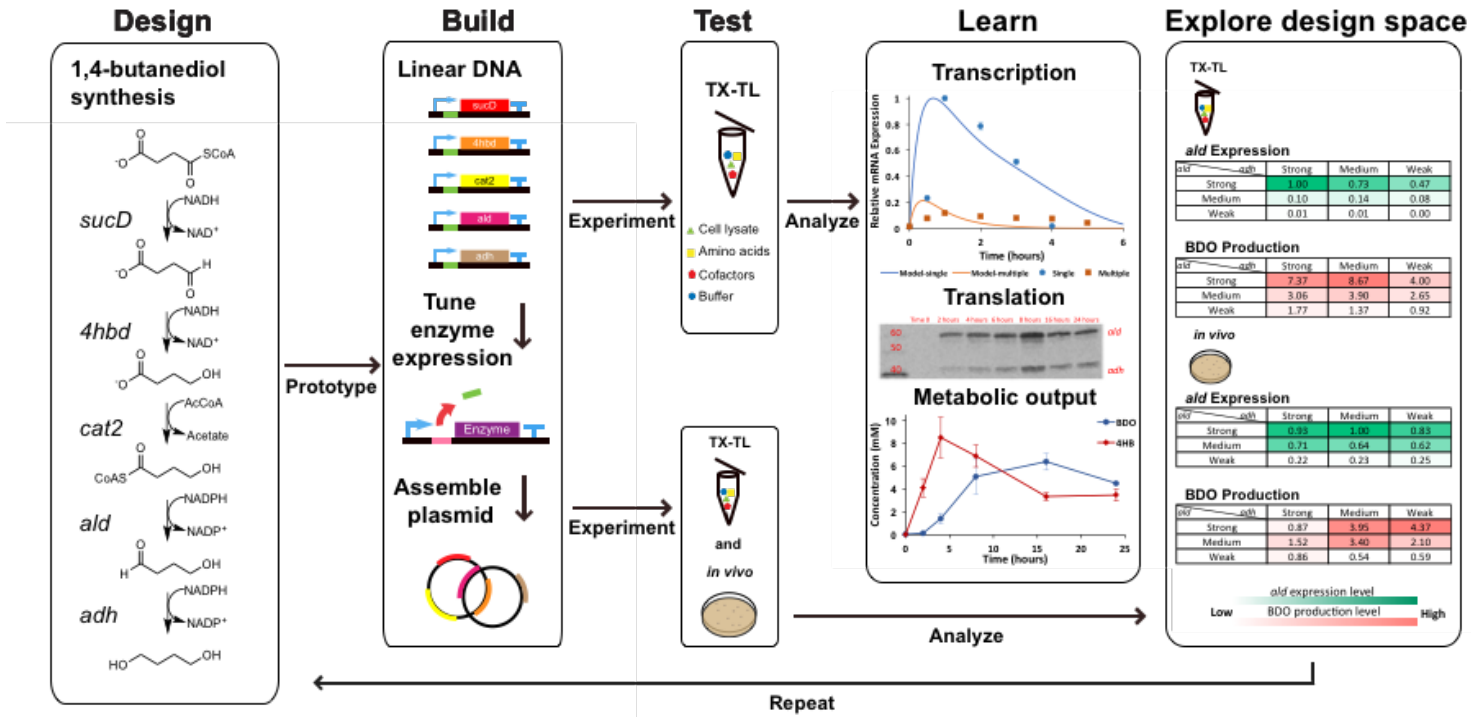


Figure 2.1: TX-TL as a platform for engineering biosynthetic pathways. A design-build-test-learn cycle using a 1,4-BDO pathway previously designed by Genomatica Inc. is shown. Linear DNA that encodes pathway enzymes is added to TX-TL. The transcription and translation machinery and metabolic output of TX-TL are analyzed. The design space of the 1,4-BDO pathway is explored in TX-TL and *in vivo* for comparison. Design: Genomatica Inc. previously designed the 1,4-BDO synthesis pathway. Build: Generate linear DNA with sequences of pathway enzymes under the same promoters, UTRs, and terminators, subsequently tune enzyme expression levels by varying UTR strength, and assemble into plasmids for design space exploration of the 1,4-BDO pathway. Test: The TX-TL system is analyzed systematically and compared to the *in vivo* system in parallel. Learn: Systematic analysis of TX-TL through reverse transcription-PCR (RT-PCR), Western blot, and metabolomics. Explore design space: Compare the expression of bottleneck enzyme and 1,4-BDO production in TX-TL and *in vivo*.

RESULTS AND DISCUSSION

Pathway Verification in TX-TL

Pathway enzymes are expressed in TX-TL reactions by adding linear DNA. Linear pieces of DNA are generated by amplifying regions encoding a promoter, a 5'-untranslated region (UTR), a coding sequence of an individual pathway enzyme, and a terminator. End products of TX-TL reactions are directly used for polyacrylamide gel electrophoresis with sodium dodecyl sulfate (SDS-PAGE) preparation, and all enzymes of the 1,4-BDO pathway show up on the gel at expected sizes (as shown in Figure S2.1). By adding linear DNA encoding enzymes from the 1,4-BDO pathway *sucD* (*035*), *4hbd* (*036*), *cat2* (*034*), *ald* (*025B*), and *adh* (*012*), we identify that the conversion from 4HB-coA to downstream metabolites is limiting the production of 1,4-BDO. After 16-hour reactions, more than 10 mM of 4HB is detected, and only 0.5 (± 0.1) mM of 1,4-BDO is detected. 3.1 (± 0.1) mM of *gamma*-butyrolactone (GBL) is also detected. GBL is the lactonized form of 4HB, and it is hypothesized to be produced spontaneously from 4HB-coA via *cat2*. The production of 4HB and GBL suggests that *sucD* (*035*), *4hbd* (*036*), and *cat2* (*034*) are not rate-limiting enzymes for the pathway, and *ald* (*025B*) and *adh (012*) can be rate-limiting. Previous results from Yim *et al.*[78] and Barton *et al.*[79] also agree with our observation of such pathway dynamics. We, therefore, hypothesize *ald* and *adh* as the bottleneck enzymes for the production of 1,4-BDO in TX-TL. To better understand TX-TL as a platform for metabolic engineering, we pick a more advanced *cat2*, *ald*, and *adh* combination, cat2 (*C*), *ald* (*C*), and *adh* (*C*), for studying a wider range of system dynamics. During an initial experiment with the advanced

enzymes, 3.4 (± 0.9) mM of 1,4-BDO is detected, and about 3.2 (± 0.5) mM of GBL is detected.

The conversion from 4HB to 1,4-BDO requires the TX-TL system to supplement for electron transfer via cofactors. We subsequently test cofactor concentrations for directing metabolic flux into the production of 1,4-BDO. We focus on cofactors directly related to the last three steps of the 1,4-BDO pathway: NADP, NADPH, acetyl-coA, and coA. The absence of acetyl-coA or coA added into the TX-TL system results in less 1,4-BDO, less GBL, but more 4HB. Also, the addition of acetyl-coA instead of coA helps produce more GBL, which translates to more 4HB-coA synthesized. We hypothesize that the TX-TL system needs more acetyl-coA for the 1,4-BDO pathway. 1 mM of acetyl-coA is added for the rest of the work. The absence of NADP or NADPH does not drastically affect target metabolite production, but the addition of NADPH improves the production of 1,4-BDO. 4 mM of NADPH is added for the rest of the work, and details can be found in Figure S2.2. These results suggest that the availability of NADPH in TX-TL is limiting the synthesis of downstream products, which indicates that more strain engineering around the TCA cycle or the pentose phosphate pathway (PPP) may help resolve cofactor imbalance in TX-TL.

System-Level Studies in TX-TL

Results from system-level studies of TX-TL system are shown in Figure 2.2. For the analysis of transcription and translation, we focus on the bottleneck enzyme *ald* (*C*). The characterization of the transcription and translation of *ald* (*C*) are carried out by adding linear

DNA encoding individual pathway enzymes. We also perform metabolomics on TX-TL reactions to grasp an understanding of the metabolic network.

Transcription

We predict and observe resource limitations of TX-TL by studying the change in mRNA expression dynamics based on how much DNA is added to the TX-TL reaction. Figure 2.2a shows the mRNA expression of *ald* (*C*) from TX-TL reaction added with just linear DNA encoding *ald* (*C*) in blue circles. The orange squares show expression of *ald* (*C*) from TX-TL reaction added with linear DNA encoding individual pathway enzymes. The blue and orange line are the respective predicted mRNA dynamics generated by the TX-TL modeling toolbox.[80] The linear DNA encoding other pathway enzymes competes with linear DNA encoding *ald* (*C*) for transcription. The mRNA expression difference of the peak mRNA value is greater than 80%. Note that 30 nM of linear DNA encoding *ald* (*C*) is added, and a total of 60 nM of linear DNA encoding other pathway enzymes is added. The transcription of *ald* (*C*) is analyzed using RT-PCR. The background signal from DNA is subtracted. The mRNA level peaks within the first hour and then drops to zero by the end of the first five hours. The rapid degradation of mRNA in TX-TL reflects resource limitations of the system. The dynamics of mRNA in TX-TL is similar to previously reported.[81]

Translation

The expression of *ald* (*C*) and *adh* (*C*) in TX-TL is analyzed using Western blots and can be normalized by total protein intensities measured from SDS-PAGE.[82] Complete SDS-PAGE and Western blots for the time-course *ald* (*C*) expression is shown in Figure S2.4. Protein

degradation is observed, which is most likely due to the presence of protease in the extract. Extract used in this work is prepared with S138 (MG 1655 △*adhE* △*ldhA* △*pflB* + *lpdA**), an engineered *E. coli* strain previously reported.[78] Most cell-free systems are either developed with purified reagents or with cell lysates from strains with protease deletion. Protein degradation is rarely captured. However, protein degradation is expected here because of the lack of protease gene knockout. The total and soluble protein expression of *ald* (*C*) and *adh* (*C*) in TX-TL are shown in Figure 2.2b. Their respective sizes are 60 kDa and 40 kDa. Most of the enzymes remain in the soluble fraction of the TX-TL reactions. The amount of each protein band on the ladder is approximately 200 ng,[83] and the concentration of *ald* (*C*) should be on the order of 10 ng/ul.

Metabolomics

To understand the metabolism in TX-TL, we carry out experiments to collect time-course data of metabolites. The production of pathway intermediate 4HB and target metabolite 1,4-BDO is shown in Figure 2.2c. The production of 4HB peaks around 4 hours into the reaction. The level of 4HB drops as the compound is converted to downstream metabolites. The concentration of 1,4-BDO increases as the concentration of 4HB decreases. The concentration of 4HB-coA peaks between 6 and 8 hours, when 1,4-BDO concentration starts to plateau, and byproducts GBL starts to accumulate (shown in Figure S2.3d). We hypothesize that the consumption of glutamate links to cofactor regeneration. The production of 1,4-BDO depends on two steps of NADPH-dependent electron transfer. The concentration of NADPH drops to the detection limit by the end of the fourth hour into the TX-TL reaction. Since NADPH is a critical cofactor for the synthesis of 1,4-BDO, the rate of NADPH being

reduced should be the same magnitude as the rate of NADPH being oxidized. The ratio of NADPH/NADP is shown in Figure S2.3b. Carbon flux through the TCA cycle is very weak, but the consumption of glutamate is very prominent. In Figure 2.2d, the conversion of glutamate reaches 80% by the end of TX-TL reactions. A preliminary ^{13}C analysis also confirms that glutamate consumption is the primary energy source (data not shown). Furthermore, the conversion from glutamate to α-ketoglutarate reduces NADP to NADPH. The amount of glutamate consumed is roughly equal to the sum of 4HB produced, GBL produced, and the 1,4-BDO produced. The consumption of glutamate is the primary metabolic flux in the TX-TL system.

We learn from the system-level studies that the first few hours are valuable for comparison between pathways or enzyme variants. TX-TL is a resource-limited system: mRNA degradation starts after the first hour, and protein degradation and cofactor imbalance happens. Metabolic flux in TX-TL mainly comes from glutamate, which is also the energy source for regenerating cofactors.

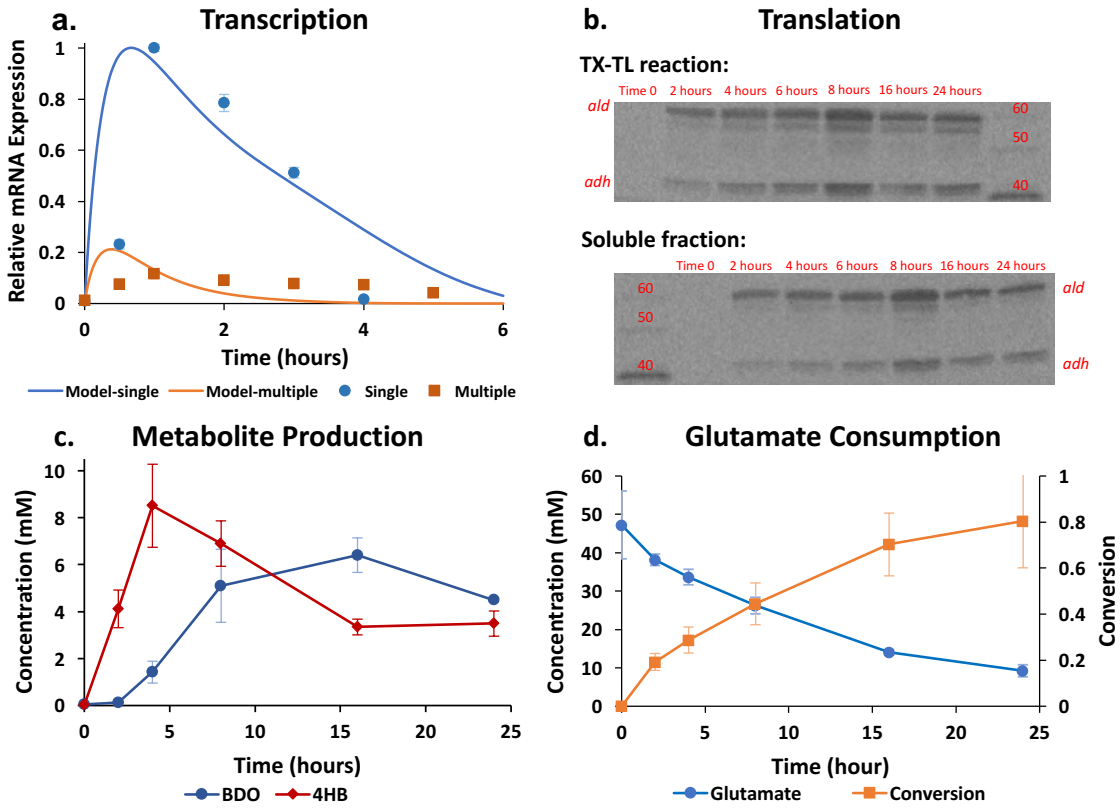


Figure 2.2: System-level studies of TX-TL. a. Transcription: Relative mRNA for *ald* (*C*) measured using RT-PCR in the first five hours of TX-TL reactions. Blue circles represent data points for mRNA expression of *ald* (*C*) from TX-TL reaction added with just linear DNA encoding *ald* (*C*), and orange squares represent data points for mRNA expression of *ald* (*C*) from TX-TL reaction added with linear DNA encoding individual pathway enzymes. The blue and orange line represents the respective predictive dynamics. b. Translation: Protein expression analyzed using Western blotting over a 24-hour time range. The top half shows results from analysis of TX-TL reactions, and the bottom half shows results from analysis of supernatant of TX-TL reactions. Both were generated from the same blot. c. Metabolite Production: Measured concentrations of 1,4-BDO (blue circle) and 4HB (red diamond) over time in TX-TL. d. Glutamate Consumption: Glutamate consumption over time in TX-TL (blue circle) and conversion (orange square). Error bars show one standard deviation for $n \geq 3$ independent experiments.

Design Space Exploration

The expression levels of *ald* (*C*) and *adh* (*C*) are modulated to explore the design space of the 1,4-BDO pathway. The expression levels of *sucD* (*035*), *4hbd* (*036*), and *cat2* (*034*) are held at fixed levels. The expression levels of all pathway enzymes are adjusted by tuning the UTR. Constructing the 1,4-BDO pathway using parts from the BCD library helps modulate pathway enzyme expression levels in TX-TL and *in vivo*. Details of construct design can be found in Figure S2.5. We generate constructs with varying enzyme expression levels, and the level of soluble *ald* (*C*) expression in TX-TL and *in vivo* are shown in Figure 2.3a and Figure 2.3d, respectively. The SDS-PAGE and Western blot images are shown in Figure S2.6. BCD 2, BCD 20, and BCD 22 are used to modify protein expression level to high, medium, and low. We have constructed plasmids with *ald* (*C*) and *adh* (*C*) using a convergent orientation to minimize genetic context effects.[84] The nine different constructs (shown in Table S2.1) show a range of protein expression levels in TX-TL and *in vivo*. The *ald* (*C*) expression levels are similar in TX-TL and *in vivo*. However, the expression of *ald* (*C*) with constructs containing BCD 22 show higher relative expression *in vivo* versus in TX-TL.

We examine the effect of the expression level of *ald* (*C*) on metabolite production in both systems. Although the metabolic output in TX-TL is not exactly the same as the one *in vivo*, we observe that the production of 1,4-BDO is closely related to the expression level of the bottleneck enzyme *ald* (*C*) in TX-TL and *in vivo,* which matches with previous studies.[70] The 1,4-BDO production level from the constructs in the two systems is different. Figure 2.3b and Figure 2.3e show the production levels of 1,4-BDO from the designed constructs.

The 1,4-BDO production from constructs with strong *adh* (*C*) expression are lower *in vivo* than in TX-TL. As shown in Figure 2.3b and Figure 2.3c, 1,4-BDO production level is linearly correlated to *ald* (*C*) expression from constructs containing the same BCD in front of the *adh* (*C*) coding region. Overall, the concentration of 1,4-BDO produced is linearly correlated to the relative expression levels of *ald* (*C*) in TX-TL regardless of the expression level of *adh* (*C*). As shown in Figure 2.3e and Figure 2.3f, the concentration of 1,4-BDO produced is linearly correlated to the relative expression levels of *ald* (*C*) *in vivo* for some constructs. The expression of *adh* (*C*) is overexpressed for constructs containing BCD 2 in front of the coding region of *adh* (*C*). Protein overexpression seems to cause a metabolic burden so strong with high *adh* (*C*) expression that the production of 1,4-BDO is limited to about 1 mM regardless of the expression level of *ald* (*C*). Data collected from both systems indicates that *ald* (*C*) is the bottleneck enzyme of the 1,4-BDO pathway.

Resource limitation is a more prominent problem *in vivo* versus TX-TL. The *in vivo* system carries a larger metabolic burden with more complicating factors such as cell growth and the development of antibiotic resistance. While developing and maintaining the transcription and translation machinery, the *in vivo* system also fights against the antibiotics in the culture broth. Although we have used promoter pA[85] as part of the constructs, leaky enzyme expression is still a problem. Transformation of constructs containing BCD 2 (for strong expression) is problematic. At 37 °C, the production of 1,4-BDO *in vivo* is much lower due to metabolic burden (data not shown). At 30 °C, the production of 1,4-BDO *in vivo* is comparable to the production of 1,4-BDO in TX-TL. The production level difference can be the supplement of NADPH and acetyl-coA in the TX-TL system.

The mapping of design space of metabolic pathway from TX-TL to *in vivo* can be further tuned by repeating the design-build-test cycle shown in Figure 2.1. There are two ways to map the two systems more closely. The first way is to redesign plasmid constructs. Re-constructing the pathway with a different set of BCDs, perhaps using BCD 9 or BCD 12 instead of BCD 2, can alleviate the metabolic burden *in vivo*. Using weaker BCDs can potentially avoid overexpression enzymes *in vivo,* which can more effectively show the linear correlation between *ald* (*C*) expression level and 1,4-BDO production. The second way is to tune protein expression in TX-TL. Plasmid concentration within cells are hard to control, but the expression level of proteins in TX-TL can be tuned by adjusting the concentration of added DNA.[60] In other words, the concentration of the added DNA can be adjusted to map protein expression levels in TX-TL to *in vivo*. In this work, we have chosen a relatively low DNA concentration to show a linear range of dynamics between *ald* (*C*) expression and 1,4-BDO production. More design iterations will lead to better mapping results.

This work has only used batch mode reaction to compare the two systems strictly. However, both systems can benefit from pH control by feeding base. Cell growth is limited by pH (shown in Figure S2.9), and the pH can be tuned by adjusting buffer.

a. ***ald* expression in TXTL**

ald \ *adh*	Strong	Medium	Weak
Strong	1.00	0.73	0.47
Medium	0.10	0.14	0.08
Weak	0.01	0.01	0.00

BDO production in TXTL

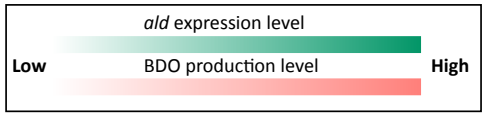
ald \ *adh*	Strong	Medium	Weak
Strong	7.37	8.67	4.00
Medium	3.06	3.90	2.65
Weak	1.77	1.37	0.92

b. ***ald* expression *in vivo***

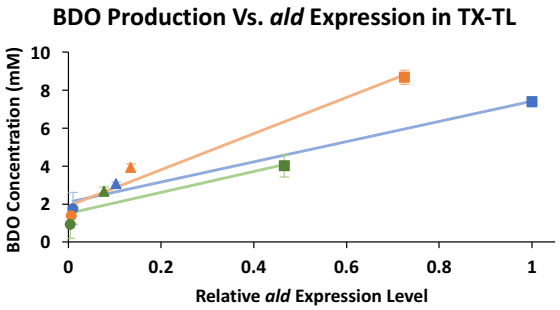
ald \ *adh*	Strong	Medium	Weak
Strong	0.93	1.00	0.83
Medium	0.71	0.64	0.62
Weak	0.22	0.23	0.25

BDO production *in vivo*

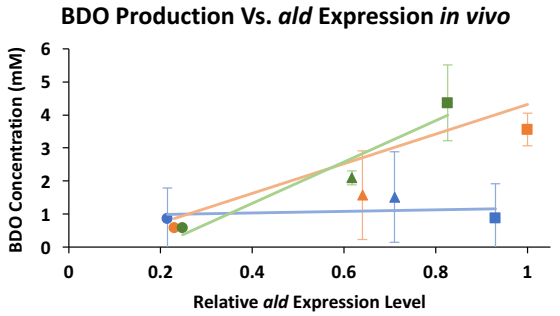
ald \ *adh*	Strong	Medium	Weak
Strong	0.87	3.95	4.37
Medium	1.52	3.40	2.10
Weak	0.86	0.54	0.59



c.



d.



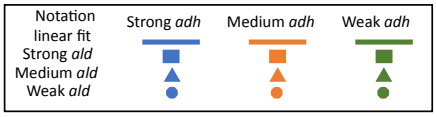


Figure 2.3: Design space exploration in TX-TL and *in vivo*. a. Heat map of *ald* (*C*) expression in TX-TL: Green—*ald* (*C*) expression as the RBS strength of the *ald* (*C*) varies. b. Heat map of the corresponding 1,4-BDO production in TX-TL: Orange—1,4-BDO production (mM) as the RBS strength of the *ald* (*C*) varies. c. 1,4-BDO production versus protein expression in TX-TL: blue—strong BCD for *adh* (*C*), orange—medium BCD for *adh* (*C*), green—weak BCD for *adh* (*C*), circle: weak BCD for *ald* (*C*), triangle: medium BCD for *ald* (*C*), square: strong BCD for *ald* (*C*). d. Heat map of *ald* (*C*) expression *in vivo*: Green—*ald* (*C*) expression as the RBS strength of the *ald* (*C*) varies. e. Heat map of the corresponding 1,4-BDO production *in vivo*: Orange--1,4-BDO production (mM) as the RBS strength of the *ald* (*C*) varies. f. 1,4-BDO production versus protein expression *in vivo:* blue—strong BCD for *adh* (*C*), orange—medium BCD for *adh* (*C*), green—weak BCD for *adh* (*C*), circle: weak BCD for *ald* (*C*), triangle: medium BCD for *ald* (*C*), square: strong BCD for *ald* (*C*). Error bars show one standard deviation for $n \geq 3$ independent experiments.

Applications of TX-TL

We apply the TX-TL system to compare three different enzyme combinations of the 1,4-BDO pathway. The reaction conditions for testing the enzyme combinations are the same. The same *sucD* (*035*) and *4hbd* (*036*) are added into TX-TL. The three different enzyme combinations are A: *cat2* (*034*), *ald* (*025B*), and *adh* (*012*) previously published,[78] B: *cat2* (*B*), *ald* (*B*), and *adh* (*B*) previously published,[79] C: evolved version of combination B, *cat2* (*C*), *ald* (*C*), *adh* (*C*). As shown in Figure 2.4, we study the production of 1,4-BDO, 4HB, and GBL in TX-TL. Combination C results with the greatest 1,4-BDO production rate, and the production of byproduct GBL is the lowest respectively. The production of 4-HB is very similar for all three combinations, but this could be that the conversion to 4-HB is reaching equilibrium by the end of the TX-TL reaction. Since the conversion from glutamate to 4-HB is redox neutral, the accumulated 4-HB is almost the same for all three combinations. Enzyme expression comparison between Combination B and Combination C is shown in Figure 2.4b. Notably, the expression of *ald* and *adh* in combination B is much weaker than the expression of *ald* and *adh* in combination C. The engineered enzymes from Combination C is evolved for stability and enzyme specificity for NADPH.

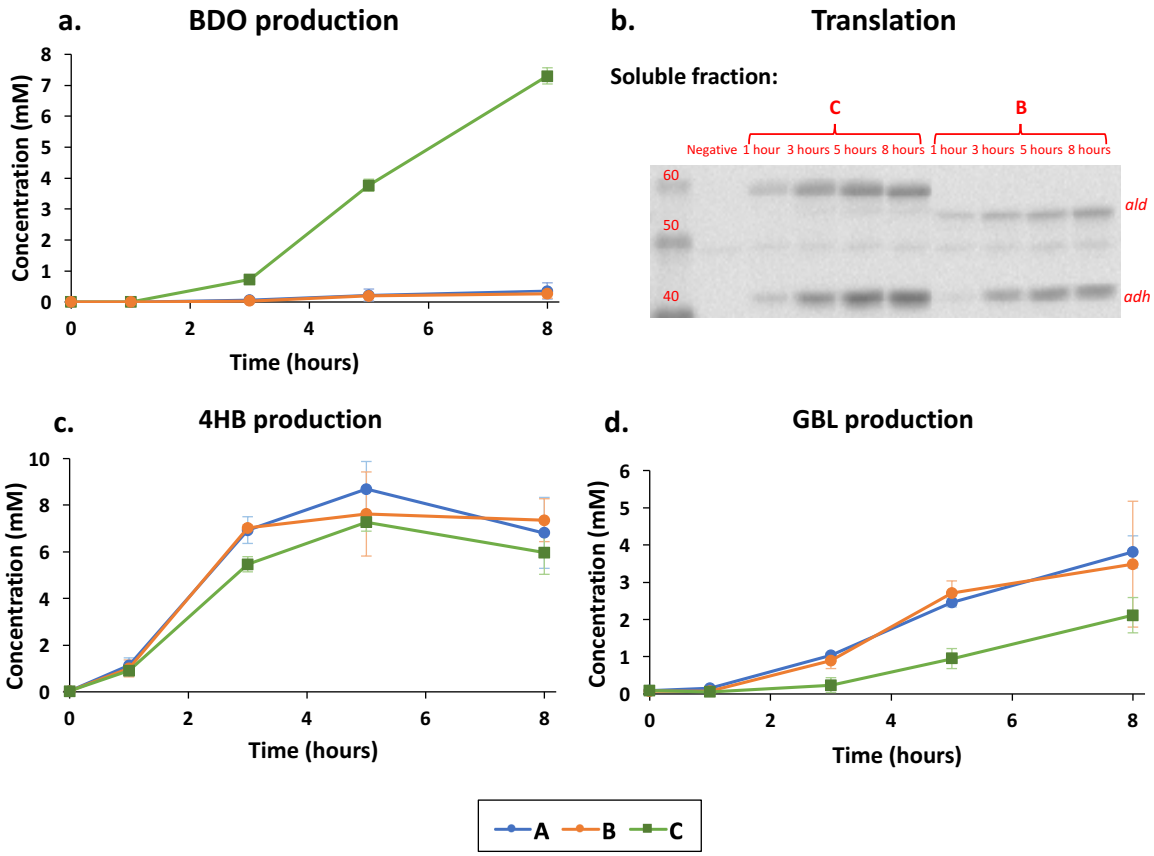


Figure 2.4: Application of TX-TL—pathway ranking. a. 1,4-BDO production in the first 8 hours of the TX-TL reaction is shown. b. Soluble protein in the first 8 hours of the TX-TL reaction is shown. c. 4HB production in the first 8 hours of the TX-TL reaction is shown. d. GBL production in the first 8 hours of the TX-TL reaction is shown. Blue circle—enzyme combination A, orange circle—enzyme combination B, green square—enzyme combination C. Error bars show one standard deviation for $n \geq 3$ independent experiments.

CONCLUSIONS

This work shows that TX-TL can be a reliable engineering platform for metabolic engineering. Through studying transcription, translation, and metabolism in TX-TL, we find that the limitation of the TX-TL system lies within rapid degradation of mRNA and cofactor imbalance. The main energy source of TX-TL is glutamate, which is different from the *in*

vivo system but this does not impact its ability to produce target metabolites similar to small-scale *in vivo* synthesis. The ease of adding DNA of interests and tuning protein expression by DNA concentrations opens up unique opportunities for rapidly exploring pathway design space. Studying the 1,4-BDO pathway in both TX-TL and *in vivo* helps confirm that TX-TL is a reliable tool for capturing pathway dynamics. Successfully demonstrating the viability of transcription and translation machinery in S138 extract and the production of 1,4-BDO provides confidence for future prototyping and engineering in extracts made with advanced *E. coli* strain or potentially other organisms.

MATERIALS AND METHODS

Cultivation

48-well plates are used for cell culture and metabolite production. M9 minimal salts medium (6.78 g l^{-1} Na_2HPO_4, 3.0 g l^{-1} KH_2PO_4, 0.5 g l^{-1} NaCl, 1.0 g l^{-1} NH_4Cl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$) is used, and it is supplemented with 10 mM $NaHCO_3$, 20 g l^{-1} D-glucose and 100 mM MOPS to improve the buffering capacity, 10 μg ml^{-1} thiamine and the appropriate antibiotics. 0.5 mM of IPTG is added to induce enzyme expression. Cell culture starts with overnight growth in LB followed by another overnight growth in minimal media and final culture. The final culture is placed on a shaker at 650 rpm at 30°C or 37°C for cultivation. The culture is centrifuged to collect pellets for protein analysis and supernatant for metabolite analysis.

Cell-Free Expression Preparation and Execution

Fermentations are performed with 1 L initial culture volume in 2-l Biostat B+ bioreactors (Sartorius Stedim Biotech). The temperature is held at 37 °C, and the pH is held at 7.0. Cells are grown aerobically to an optical density (O.D.) of approximately 10, at which point the culture goes through spin-down, re-suspense, and wash cycle series, according to protocol previously described by colleagues.[33] Collected cell pellets are homogenized using M-110F Microfluidizer Processor and extracted according to the method described by Kwon *et al.*[86] with the post-homogenization incubation period extended to 80 min instead of 60 min. Extract activity is verified initially and tested for optimal salt concentration using GFP. The concentration of potassium glutamate is capped at 30 mM to capture metabolic dynamics of the system. Buffer preparation is done according to protocol developed from previous studies[33] with a supplement of 30 mM maltodextrin. The preparation results in extract with conditions: 8.9−9.9 mg/mL protein, 4.5−10.5 mM Mg-glutamate, 20-40 mM K-glutamate, 0.33−3.33 mM DTT, 1.5 mM each amino acid except leucine, 1.25 mM leucine, 50 mM HEPES, 1.5 mM ATP and GTP, 0.9 mM CTP and UTP, 0.2 mg/mL tRNA, 0.26 mM CoA, 0.33 mM NAD+, 0.75 mM cAMP, 0.068 mM folinic acid, 1 mM spermidine, 30 mM 3-PGA, 4 mM NADPH, 1 mM acetyl-coA, and 1 mM NADH. When needed, inducers such as IPTG, linear DNA, or plasmid DNA are added to a mix of extract and buffer. TX-TL reactions are conducted in PCR tubes and kept at 29°C with incubation in PCR machine. BioTeK Synergy H1 microplate reader is used to collect kinetic data for fluorescent protein and MG-Aptamer. Protein gamS is added into TX-TL reactions to prevent linear DNA degradation.[60]

Plasmid DNA, PCR Product Preparation, and Cloning

PCR products are amplified using KOD (Novagen, EMD). Plasmids are miniprepped using a QIAprep spin columns (Qiagen) and further concentrated with Millipore's Amicon Ultra-0.5 Centrifugal Filter Unit with Ultracel-30 membrane. All plasmids are processed at the stationery phase. Before use in the cell-free reaction, PCR products undergo an additional PCR purification step using QIAquick PCR Purification Kit (Qiagen), which removes excess salt detrimental to TX-TL and are eluted and stored in water at −20°C for long-term storage. Gibson Assembly Ultra and Gibson Assembly HiFi Master Mix from SGI-DNA are used for plasmid assembly.

Analytical Methods

LCMS/MS is used to analyze TX-TL samples and supernatant of cell culture. TX-TL samples were first diluted with 1:3 volume ratio with methanol to remove proteins and other big molecules, and then diluted 1: 12.5 with diluent containing labelled internal standards before LCMS analysis. Cell culture samples were diluted 1:50 with same diluent before LCMS analysis. API3200 triple quadrupole system (AB Sciex, Life Technologies, Carlsbad, CA), interfaced with Agilent 1260 HPLC, utilizing electrospray ionization and MRM based acquisition methods is used. 1,4-BDO and GBL are detected using positive ionization mode, while 4HB and related acidic compounds are detected using negative ionization mode. Zorbax Eclipse XDB C18 4.6x30mm (particle size 1.8um) was used. Column temperature is maintained at 40°C, flow rate of 0.7 mL/min. Injection volume is 5 ul. Eluents include water

with 0.1% formic acid (A) and methanol with 0.1% formic acid (B). A fast 1.5 min 5-95% methanol gradient is used, resulting in 1.5 min long LCMS method.

Protein Gel and Western blotting

Culture pellets are re-suspended with Novagen BugBuster Protein Extraction solution mixed with rLysozyme and benzonase. The amount of BugBuster added is normalized based on the OD value of 1mL culture. Samples are incubated at room temperature for 20 minutes. An equal volume of 2X Laemlii buffer containing 5% of ß-mercaptoethanol is mixed with each sample. The mixture is subsequently boiled in a thermocycler at 99°C for 5 min. 10ul of a mixture containing cell pellets is loaded into each lane, or 5 ul of a mixture containing TX-TL sample is loaded. Invitrogen MagicMarker XP Western Protein Standard and Bio-Rad Kaleidoscope is added as protein ladders. Bio-Rad 4-15% Criterion TGX (Tris-Glycine eXtended) Stain-Free precast gels are used. Gel imaging is done using Bio-Rad Gel Doc EZ Imager for measuring total protein intensity. Invitrogen iblot system is used for gel transfer. Western blot imaging is done using the Chemi Hi Resolution setting on a BioRad ChemiDoc MP imager. The intensity from Western blot is normalized by the total protein from each lane, and stain-free total protein is the loading control for Western Blots.[82] Protein band intensity is determined by using Image Lab 5.2.1.

COMPETING FINANCIAL INTERESTS

Support for this work is provided by Genomatica Inc., a company pursuing commercialization of the 1,4-butanediol process discussed here. All authors except Y.Y.W. and R.M.M. are employees of Genomatica Inc. at the time the work is performed. R.M.M is a co-founder and board member of Synvitrobio, a for-profit company developing cell-free systems.

ACKNOWLEDGEMENTS

We thank Dr. Nathan Dalleska and the Environmental Analysis Center for the support and assistance using GC/MS for initial data collection. We thank Anna Lechner, Evan Ehrich, Emily Mitchell, Glenn Majer, and Sari Rizek for material analysis and preparation. We thank Jingyi Li, Jungik Choi, Jonathan Joaquin, Joseph R. Warner, Robin Osterhout, and Robert Haselbeck for helpful discussion. This material is based upon work supported in part by the Defense Advanced Research Projects Agency (DARPA/MTO) Living Foundries program; contract number HR0011-12-C-0065 (DARPA/CMO). Y.Y.W was supported by NIH/NRSA Training Grant 5 T32 GM07616 and the Gordon and Betty Moore Foundation Grant GBMF2809 to the Caltech Programmable Molecular Technology Initiative.

SUPPLEMENTARY INFORMATION

Pathway Verification in TX-TL

Linear DNA encoding individual enzymes are added into TX-TL reactions to verify enzyme expression. End products of TX-TL reactions are directly used for polyacrylamide gel electrophoresis with sodium dodecyl sulfate (SDS-PAGE) preparation, and all enzymes of the 1,4-BDO pathway show up on the gel at expected sizes. Below Figure S2.1 shows an SDS-PAGE gel with the expression of relevant pathway enzymes. 0.5 ul of TX-TL reactions are used for analyzing enzyme expression. Background proteins are proteins extracted during cell lysis.

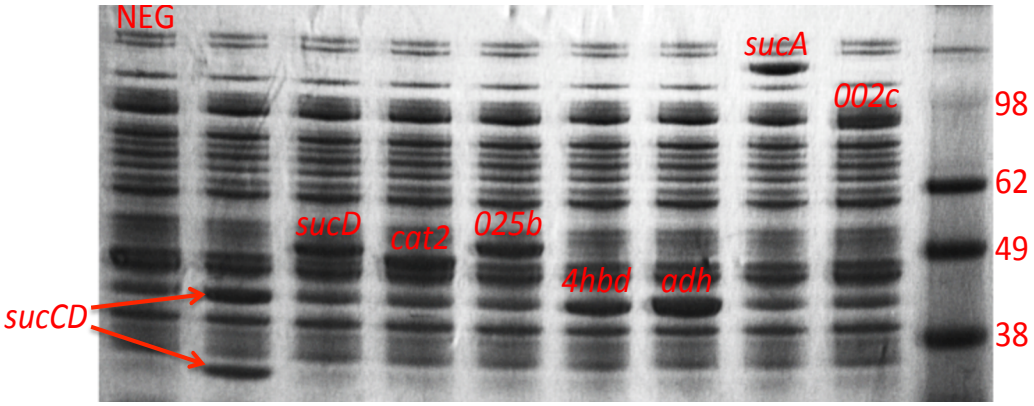


Figure S2.1: Verifying enzyme expression using SDS-PAGE gel: From left to right are negative control (NEG), succinyl-CoA synthetase (*sucCD*, two subunits α and β at 29.6 kDa and 41.4 kDa[87]), CoA-dependent succinate semialdehyde dehydrogenase (*sucD-035*, 50.2 kDa), 4-hydroxybutyryl-CoA transferase (*cat2-034*, 48.0 kDa), 4-hydroxybutyryl-CoA reductase (*ald-025B*, 52.1 kDa), 4-hydroxybutyrate dehydrogenase (*4hbd-036*, 41.3 kDa) alcohol dehydrogenase (*adh-012*, 43.1 kDa), 2-oxoglutarate decarboxylase (*sucA*, 100kDa), 4-hydroxybutyryl-CoA reductase (*ald(002c)*, 95.3kDa), and protein ladder.

A range of cofactor concentration were tested for improved 1,4-BDO production. Results are shown in Figure S2.2. In a typical TX-TL reaction, 0.26 mM of coA has been added as part of the buffer, 1 mM of NADPH, 1 mM of acetyl-coA, and 1 mM of NADH are added. Here, we vary the concentration of coA, acetyl-coA, NADP, and NADPH one at a time to understand the effect of these cofactors have on the metabolite production. The metric here is to compare downstream metabolite production versus upstream metabolite production. Mainly, we are comparing the total production of 1,4-BDO and GBL to the total production of GABA and 4HB. Figure S2.2 shows that the lack of coA or acetyl-coA leads to the accumulation of upstream metabolites. When more NADPH is added to the system, more downstream metabolites are produced.

System-level Studies of TX-TL

The production of acetate is very prominent in the TX-TL system. The production of acetate through the 1,4-BDO pathway enzymes does not account for the acetate measured in the system. The conversion from 3-PGA or glutamate to acetate regenerates cofactor such as NADH, NADPH, ATP, and acetyl-coA, which explains the significant production of acetate.[88] Figure S2.3b and Figure S2.3c below shows the ratio of NADPH/NADP and NADH/NAD. The energy charge is also calculated to capture resource limitation of the system (shown in Figure S2.3a). The production of HB-coA in Figure S2.3d shows the peak at 6 hours, which is when 1,4-BDO production starts to plateau.

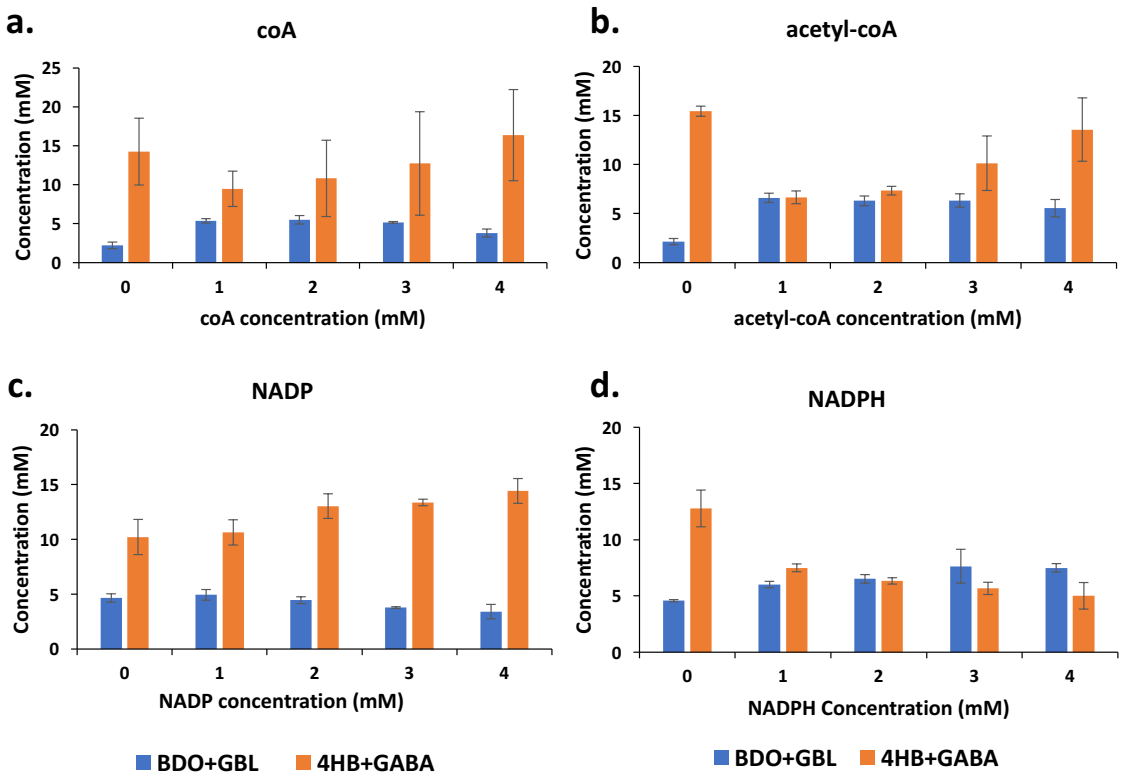


Figure S2.2: Testing cofactor concentration for improved 1,4-BDO production: a. coA b. acetyl-coA c. NADP d. NADPH, blue bar—sum of the concentration of 1,4-BDO and GBL, orange bar—sum of the concentration of 4HB and γ-aminobutyric acid (GABA). Error bars show one standard deviation for $n \geq 3$ independent experiments.

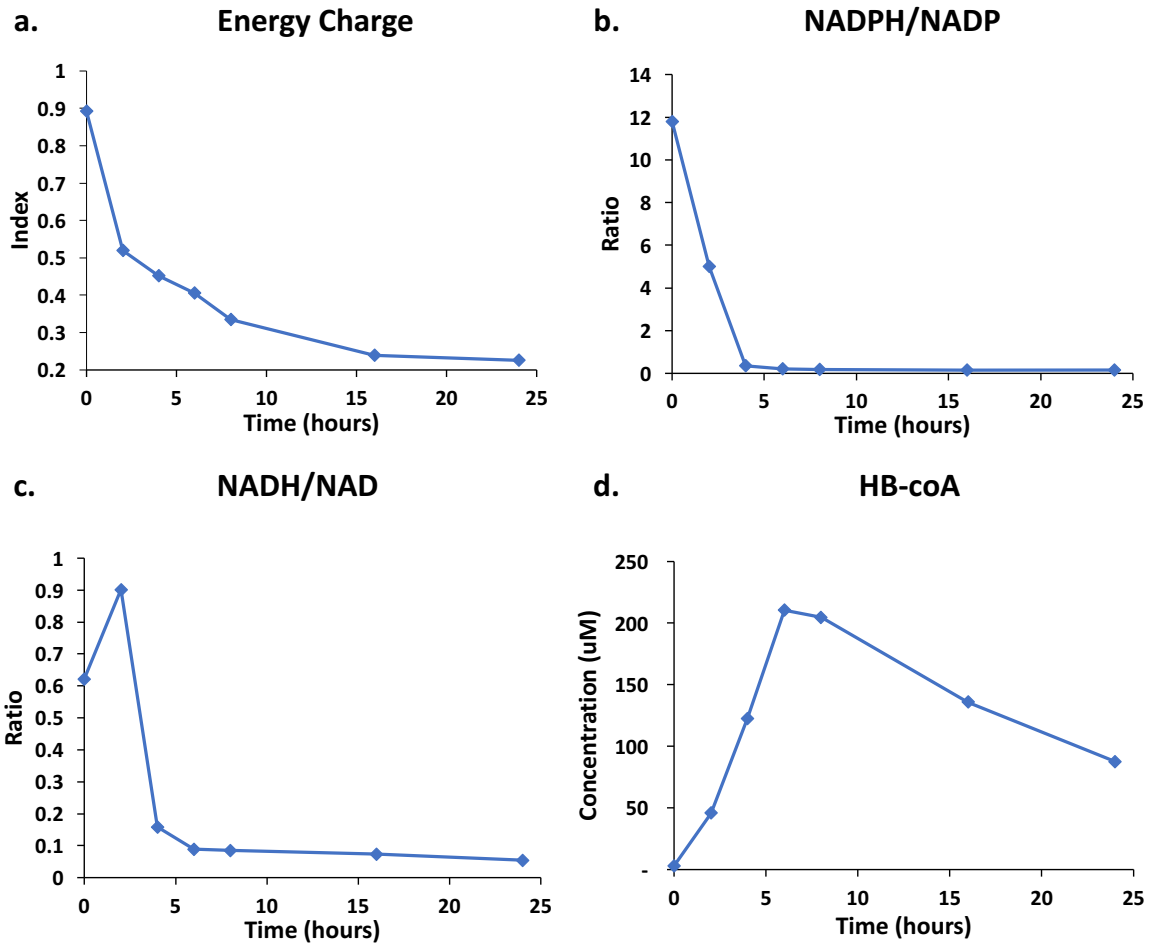


Figure S2.3: Metabolomics in TX-TL: a. Energy charge over time b. Ratio of NADPH/NADP c. Ratio of NADH/NAD d. HB-coA concentration over time.

Genomatica has previously developed antibodies specifically for enzyme *ald* and *adh*. Without inserting additional protein tags, we analyze enzyme expression profile without interfering pathway dynamics. Below in Figure S2.4 shows the SDS-PAGE (top) and Western blot (bottom) that analyzed the protein dynamics over time. The protein bands in the Western blot show unspecific binding at 30 kDa and 80 kDa. However, the separation between target protein band and the unspecific band is large enough for protein expression analysis. As shown, the *ald* (*C*) is about 60 kDa, where the *adh* (*C*) is about 40 kDa. We

analyzed the same sample in two different ways: 1) add TX-TL reaction directly for protein gel analysis and 2) centrifuge samples and add soluble fraction for protein gel analysis. The protein analysis shows that TX-TL is a good expression system where a good amount of protein remains in the soluble fraction. Here we can also visually identify that protein expression peaks at 8 hours.

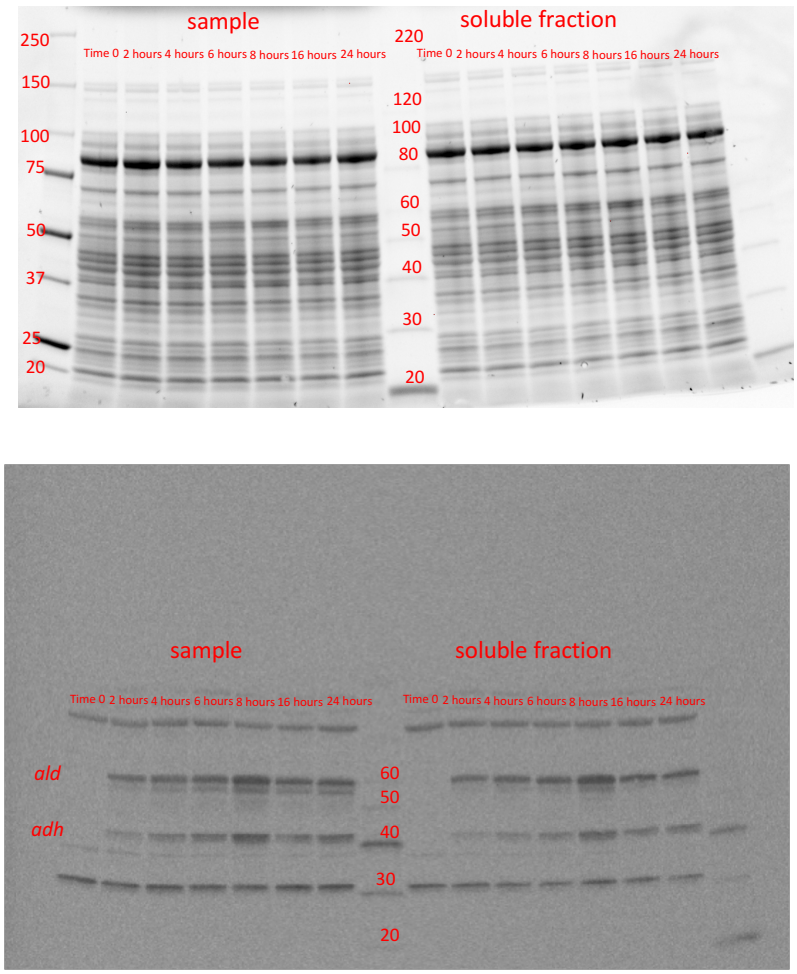


Figure S2.4: *ald* (*C*) and *adh* (*C*) expression over time in TX-TL. Top: SDS-PAGE gel. Bottom: Western blot. Both images are separated by total sample (left) and soluble protein fraction (right). *ald* (*C*) and *adh* (*C*) are 60kDa and 40 kDa respectively.

Design Space Exploration

We have picked elements from Chen *et al.*'s terminator library:[89] ECK120033736 (164.6x), ECK120015440 (119.21x), and ECK120010799 (101.05x). These terminators have been previously tested for similar expression levels. We use plasmids pZ33s and pZS*23s previously published.[78] Detailed constructs are shown in Figure S2.5. We use BCD22 for enzymes on the pZA33s plasmid. BCD 22 is the UTR that gives weak protein expression, and pZA33s has p15A as a replication of origin. pZS*23S has SC101 as a replication of origin. The pZA33s plasmid has a slightly higher copy number than the pZS*23S. For experiments done during design space exploration, 5 nM of the pZA33s is added, and 3 nM of pZS*23s is added. The ratio used in TX-TL is an attempt to reflect the copy number ratio *in vivo*. PZA33s contains kanamycin resistance, and pZS*23s contains chloramphenicol resistance. While kanamycin kills bacteria by binding to 30S ribosomal subunit and causes misreading of t-RNA, chloramphenicol kills bacteria by binding to 50S ribosomal subunit and prevents peptide bond formation.[90] To better map TX-TL to *in vivo*, it may also be a good idea to switch to a different set of antibiotics.

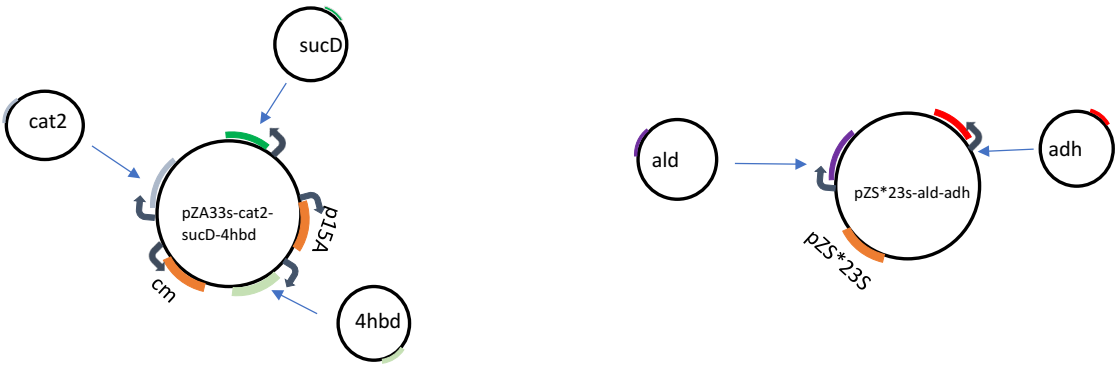


Figure S2.5: Construct assembly for design space exploration. Left: non-bottleneck enzymes *sucD*, *cat2*, and *4hbd* on pZA33s. Right: bottleneck enzymes *ald* and *adh* on pZS*23s.

Constructs are designed to mimic the ratio of linear DNA encoding pathway enzymes. The metabolic profile of TX-TL reaction added with plasmid DNA is similar to the metabolic profile of TX-TL reaction added with linear DNA. Table S2.1 shows the constructs with varying *ald* (*C*) and *adh* (*C*) strength. The numbers in the table are the order of which the samples are loaded onto the protein gel for expression comparison between TX-TL and *in vivo*.

Sample Order		*adh* Expression Level		
		Strong	Medium	Weak
ald Expression Level	Strong	1	2	3
	Medium	4	5	6
	Weak	7	8	9

Table S2.1: Constructs with varying *ald* (*C*) and *adh* (*C*) expression levels: The numbers in the table are the order in which samples are loaded onto protein gels for expression comparison shown in Figure S2.7.

Samples for metabolite analysis and protein expression analysis are collected after 16 hours of TX-TL reaction. The raw SDS-PAGE and Western blots are shown in Figure S2.6. Protein ladder followed by the negative control are added as the two samples. Constructs with varying *adh*(*C*) and *ald* (*C*) expression levels shown in Table S2.1 then added by its numerical order. Values presented in Figure 2.3 are intensities of soluble protein from the Western blot, which are measured and normalized by the total protein intensities shown in the SDS-PAGE gel. Overall, Figure S2.6b and Figure S2.6d show a range of protein expression levels for *adh* (*C*) and *ald* (*C*). Note that there are more unspecific bands in the

Western blot for *in vivo* samples. We have observed insoluble proteins in the *in vivo* samples, which explains the smear on the right side of the Western blot shown in Figure S2.6d. Table S2.2 lists the parameters from linear regression for data shown in Figure 2.3c and Figure 2.3f. The slope differences in TX-TL indicates that resource limitation might have played a role in metabolite production. The strong expression of *adh* (*C*) may have impacted metabolite production, where too weak of an expression is not enough for metabolite production. The slope of the linear regression for *in vivo* data is very similar, which confirms that *ald* (*C*) is the bottleneck enzyme for the production of 1,4-BDO.

Linear Regression Parameters in Figure 2.3c		
adh Expression	slope	intercept
Strong	5.32	2.09
Medium	9.52	1.90
Weak	5.52	1.50

Linear Regression Parameters in Figure 2.3f		
adh Expression	slope	intercept
Strong	0.24	0.94
Medium	4.49	-0.17
Weak	6.26	-1.18

Table S2.2: Slope and intercept for linear regression in Figure 2.3.

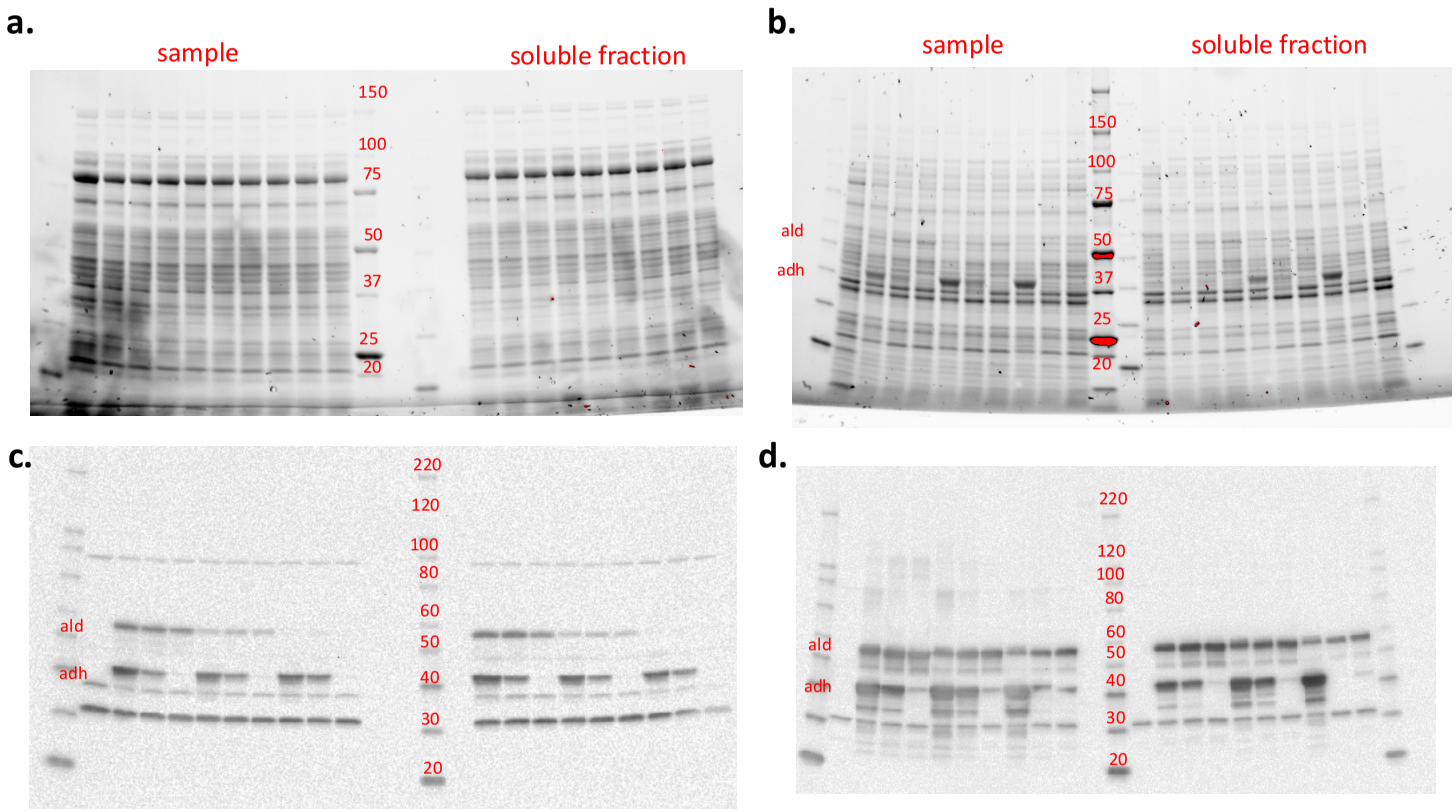


Figure S2.6: *ald* (*C*) and *adh* (*C*) expression *in vivo* (right: b and d) Vs. in TX-TL (left: a and c). Top: SDS-PAGE gel (a and b). Bottom: Western blot (c and d). Each image is separated by total sample (left) and soluble protein fraction (right). *ald* (*C*) and *adh* (*C*) are 60kDa and 40 kDa respectively.

Applications of TX-TL

Enzyme expressions from two of the three enzyme combinations are analyzed. Details are shown in Figure S2.7. Time-course protein expression (1 hour, 3hours, 5 hours, and 8 hours) of Combination B and Combination C are analyzed. From left to right of the Western blot, a negative control was added first, and time-course samples from Combination B and Combination C were added. To the left of the Western blot is the relative *ald* expression over time. Note that enzymes from Combination C have a strong expression than enzymes from Combination B. Both *ald* and *adh* from Combination C are enzymes engineered for stability *in vitro*, while the ones from Combination B are wild-type enzymes.

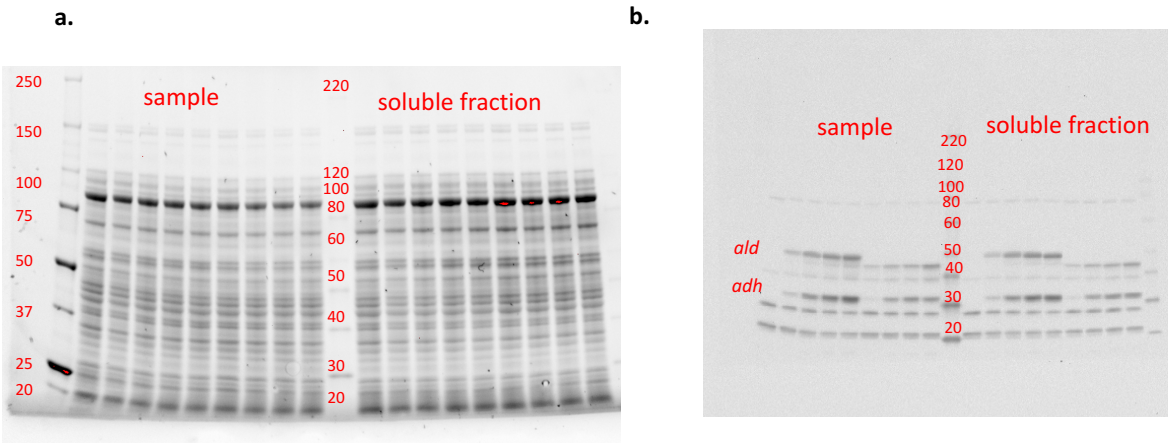


Figure S2.7: *ald* and *adh* expression for ranking pathways in TX-TL. To the left of the ladder are samples from total TX-TL reactions, and to the right of the ladder is soluble protein in the supernatant fraction of TX-TL reactions.

Characteristics of S138 *in vivo*

Metabolomics of S138 *in vivo* is also performed. Below in Figure S2.8 shows the production of 1,4-BDO over time. Energy charge (Figure S2.8b), NADPH/NADP ratio (FigureS2.8c),

and NADH/NAD ratio (Figure S2.8d) are very similar to the ones observed in TX-TL. The pH *in vivo* drops rapidly (Figure S2.9b), and it is about 4 by the end of 28 hours, where the pH in TX-TL stays above 6 during 8-hour reactions. Improvement on the buffer of the *in vivo* system can help improve metabolite production.

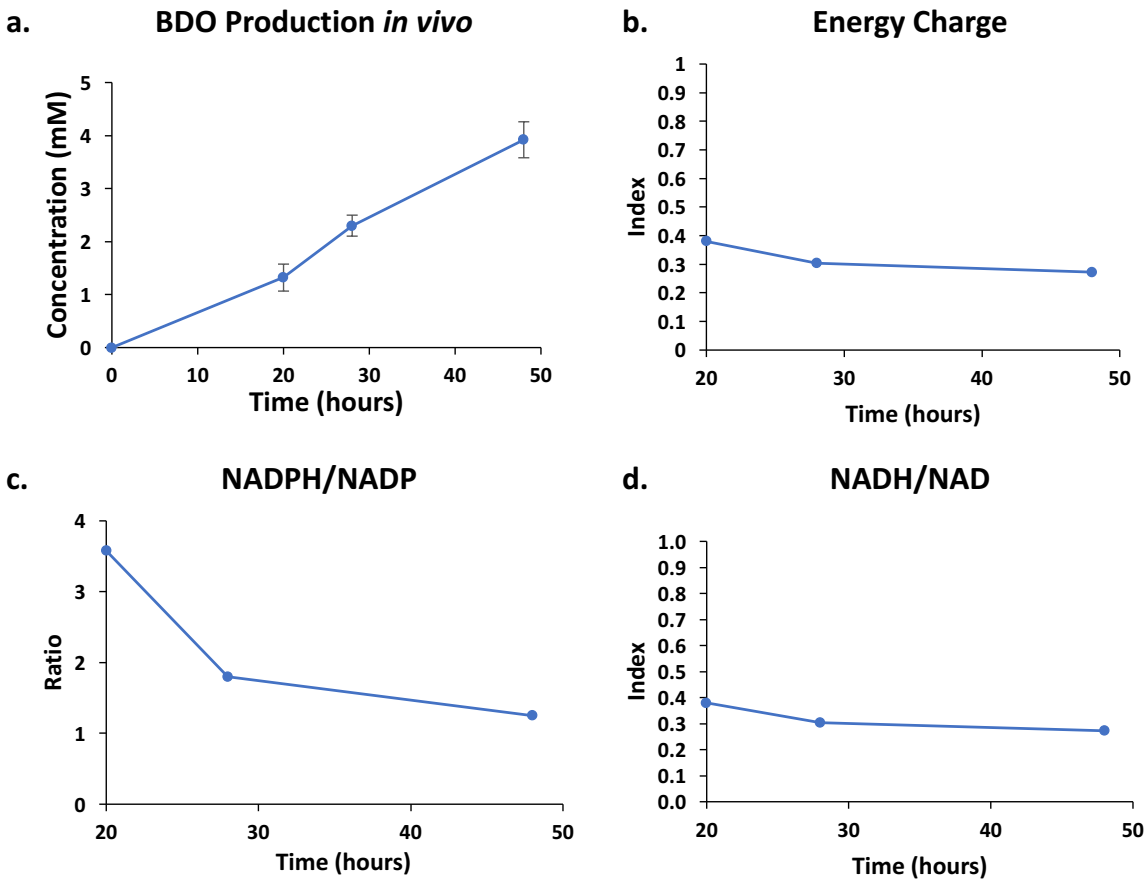


Figure S2.8: Characteristics of S138 *in vivo*: a. 1,4-BDO production over time b. pH over time c. Energy charge over time, d. NADPH/NADP and e. NADH/NAD. For c, d, and e: blue—negative control, orange—S138 with the 1,4-BDO pathway (Data are collected with Construct 4 shown in Table S1).

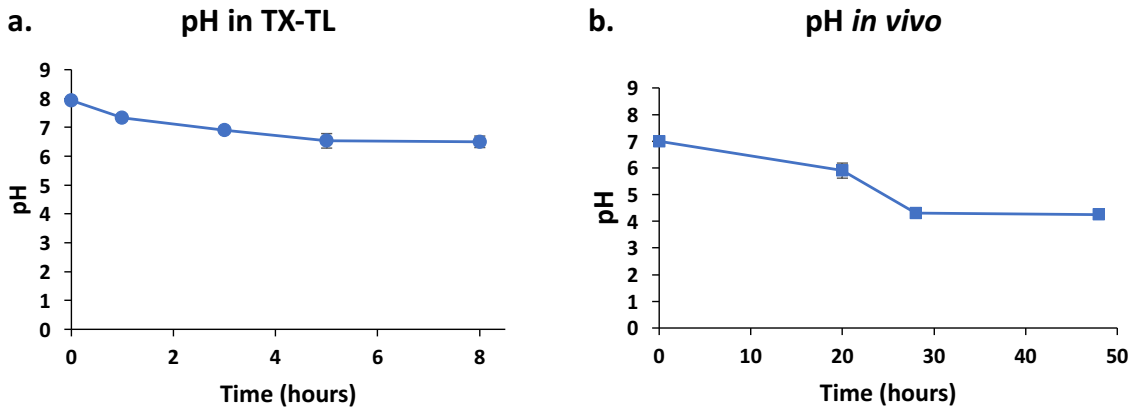


Figure S2.9: pH comparison between TX-TL and *in vivo*: a. pH measured over time of TX-TL samples b. pH measured over time of *in vivo* samples.

Chapter 3

DIVERSITY-ORIENTED BIOSYNTHESIS IN CELL-FREE SYSTEMS: EXPANDING SUBSTRATE SCOPE OF THE VIOLACEIN PATHWAY IN TX-TL

Authors: Yong Y. Wu, David Romney, Joy Doong, Frances H. Arnold, Richard M. Murray

Keywords: violacein, TX-TL, drug discovery, tryptophan, cell-free biosynthesis, substrate scope

ABSTRACT

The cell-free transcription-translation (TX-TL) system provides a reliable platform for characterizing and engineering a biosynthetic pathway. TX-TL simulates a customizable cellular environment that can be controlled by manipulating pH, changing cellular components, or adding exogenous substrates. By adding linear DNA encoding individual enzymes of the violacein pathway and tryptophan analogs in TX-TL reactions, we have discovered new violacein analogs. TX-TL enables rapid production of natural product analogs with diverse substitution, which allows small-scale biosynthesis of potential drug candidates and offers a new platform for drug discovery. This work also presents TX-TL as a platform for protein engineering. Residues targeted for site-saturated mutagenesis were identified with protein-ligand docking. Linear DNAs of individual enzyme mutants were added into TX-TL reactions to screen for improved enzyme variant. Screening result indicates vioE mutant Y17H reduces byproduct formation and redirects metabolic flux towards target metabolites. Protein engineering for improved enzyme activity can further expand the substrate scope of a natural product pathway and result with more natural product analogs that can be applied for medical applications.

INTRODUCTION

Engineering proteins for novel chemistry has been one of the main pillars of biosynthesis. Recent studies have shown the directed evolution of a cytochrome c for carbon-silicon bond formation,[91] the engineering of P450 for the biosynthesis of an intermediate for the production of ticagrelor,[92] and the engineering of TrpB for tryptophan analogs.[56] The application of protein engineering for diversity-oriented biosynthesis can significantly impact drug discovery. Natural products have the potential to become antibiotics and therapeutics.[43] Further, natural product analogs have previously shown to have increased antibacterial activity when compared to the natural products.[93] The addition of functional groups such as methyl, fluoro, and nitrile to natural products can improve compound rigidity,[48] modulate cation-π interaction,[49] and polarize adjacent electron density.[50] Such modifications can enhance the potency, lipophilicity, bioavailability, and metabolic stability of a drug candidate.[51-52] Engineering pathway enzymes for the production of natural product analogs will expand the chemical space for biosynthesis and drug discovery.

This work aims to apply protein engineering for expanding the substrate scope of the violacein pathway in the cell-free transcription-translation (TX-TL) system (shown in Figure 3.1). This work demonstrates the viability of using linear DNA in the cell-free TX-TL system[60] to accelerate the engineering of biosynthesis for natural product analogs. The TX-TL system offers direct access to cellular biochemistry and provides a robust environment for exogenous protein synthesis and metabolite production. Using the TX-TL system as a screen for protein engineering reduces the difficulty of optimizing buffer conditions and

tackling growth-coupled challenges, such as metabolic burden and protein toxicity. The use of TX-TL for protein engineering also allows the substrate scope of complex pathways to be surveyed rapidly, simplifies efforts for pathway construction and enzyme expression tuning.

This work showcases the engineering of the pathway to produce violacein, a natural product found in *Chromobacterium violaceum*.[94] Studies have shown that violacein and deoxyviolacein have anti-cancer,[95-96] antiparasitic,[97] and other valuable properties.[98-100] The violacein pathway (Figure 3.2) was previously characterized *in vitro* by Balibar *et al.*,[101] and the biosynthesis of violacein in cell-free systems has been reported.[62, 102-103] Initially, enzyme vioA converts tryptophan (**1a**) to indole-3-pyruvate (IPA, **2a**), which is then converted by vioB to intermediate (**3a**). This intermediate is unstable, and can spontaneously form chromopyrrolic acid (CPA, **4a**) through a single-electron pathway initiated by radicals generated from oxygen and reduced flavin. Alternatively, enzyme vioE can intercept **3a** and convert it to prodeoxyviolacein (**5a**). Finally, mono-oxygenase vioD converts **5a** to proviolacein (**6a**), which is then converted to violacein (**8a**) by oxidase vioC. Intermediate **5a** is also a competent substrate for vioC, in which case deoxyviolacein (**7a**) is the final product.

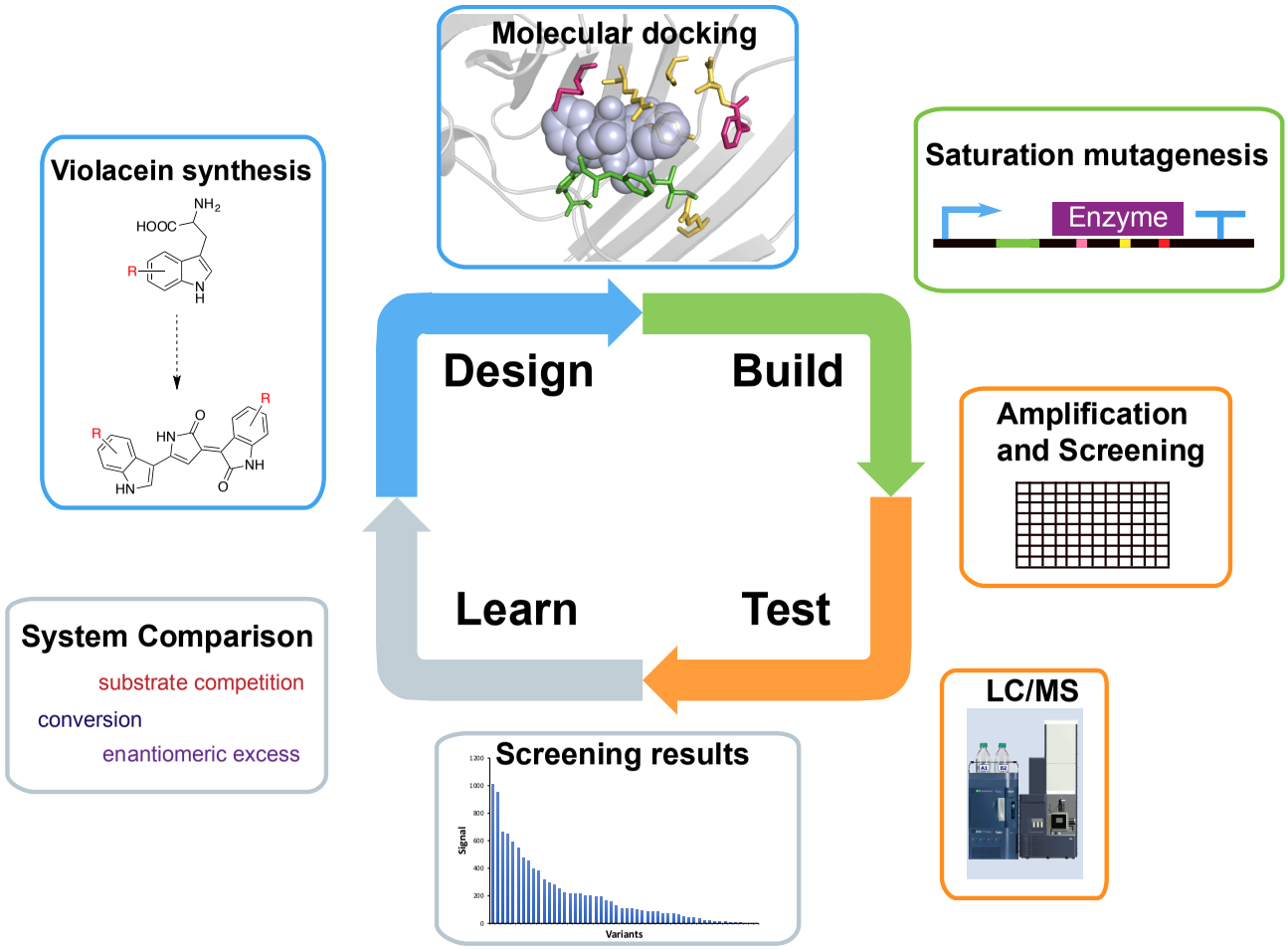


Figure 3.1: A biochemical platform to expand pathway substrate scope.

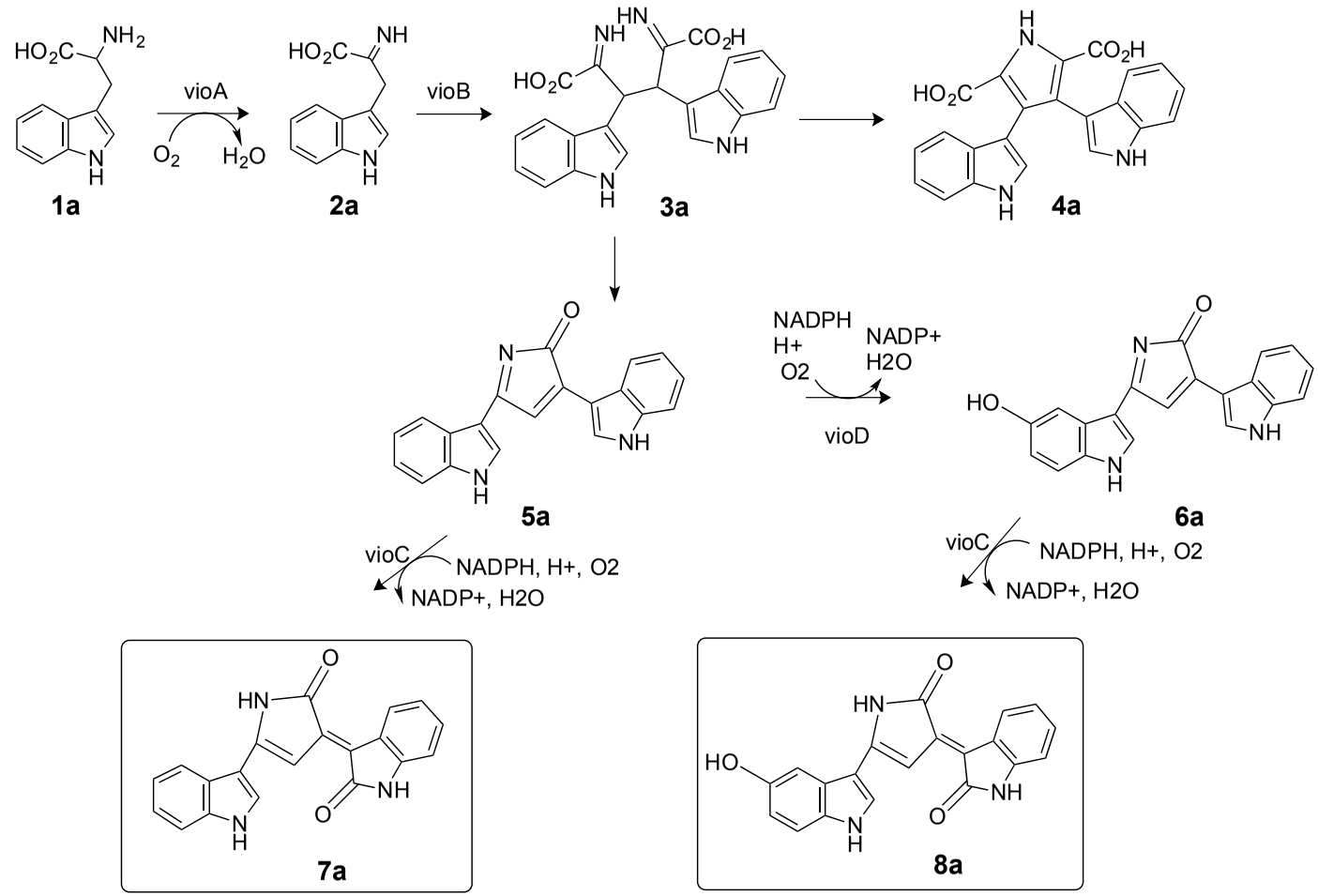


Figure 3.2: The biosynthetic pathway for violacein.

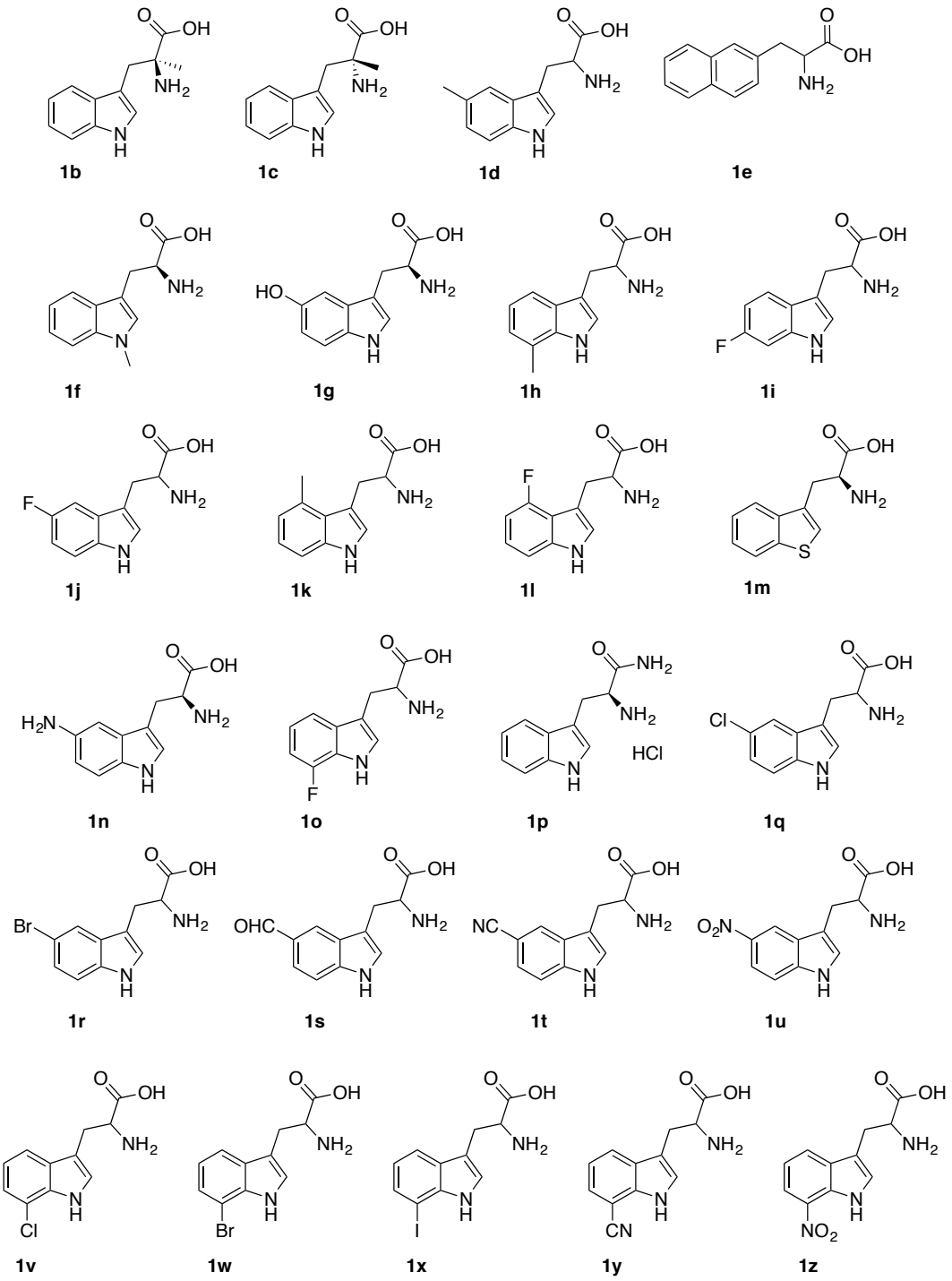


Figure 3.3: Tryptophan analogs for substrate scope studies.

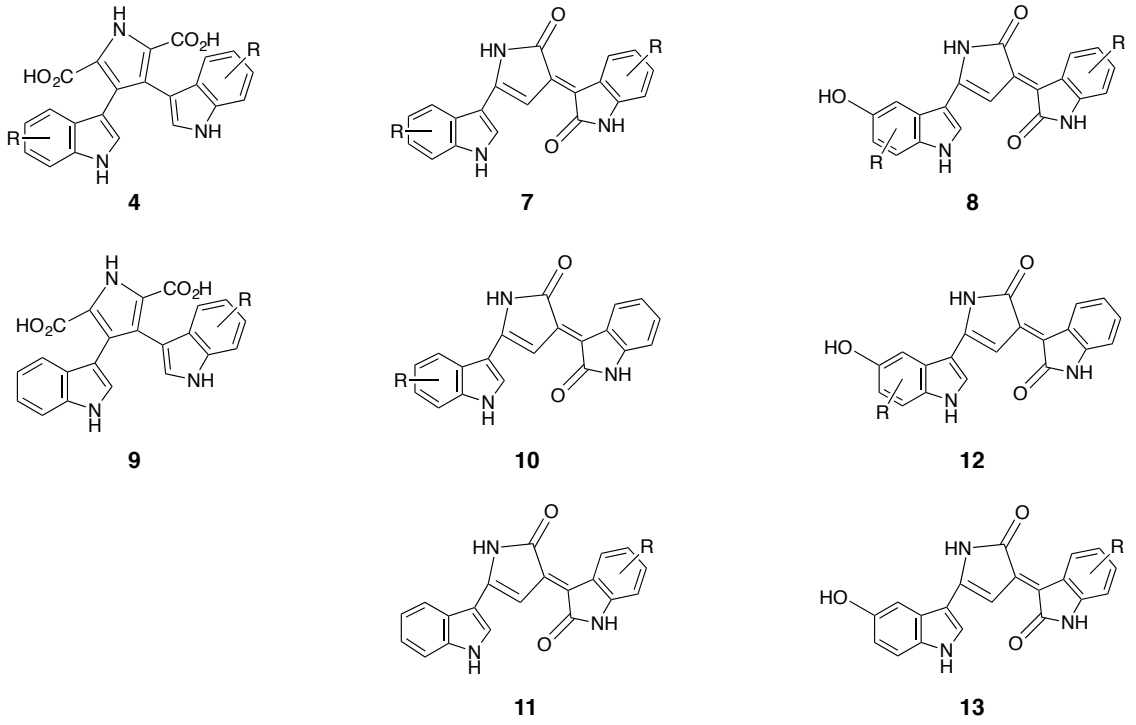


Figure 3.4: Detectable products from the violacein pathway.

RESULTS AND DISCUSSION

Twenty-five tryptophan analogs (**1b** to **1z,** Figure 3.3) with diverse substitution patterns were used to probe pathway enzyme promiscuity of the pathway enzymes. Figure S3.1 shows the number of the positions on the indole moiety of tryptophan. For the rest of the discussion, the alphabetical designation for each tryptophan analog also applies to the associated products. Figure 3.4 shows detectable downstream metabolites of the pathway, which are from left to right: homodimers of CPA analogs (**4** or CPA_RR), heterodimers of CPA analogs (**9** or CPA_R), homodimers of deoxyviolacein analogs (**7** or deoxyviolacein_RR), heterodimers of deoxyviolacein analogs (**10** and **11** or deoxyviolacein_R), homodimers of

violacein analogs (**8** or violacein_RR), and heterodimers of violacein analogs (**12** and **13** or violacein_R).

Substrate Scope Studies

Tryptophan analogs **1b** to **1p** were added into TX-TL reactions to study substrate scope of the violacein pathway, and results are consistent with the mechanism previously proposed.[101] Preliminary screening used TX-TL to test for the production of deoxyviolacein (**7**,**10**, or **11**) by adding linear DNA encoding vioA, vioB, vioE, and vioC (Figure 3.5). Wild-type enzymes from the violacein pathway tolerated moderate decoration of the indole ring. Tryptophan analogs with the indole moiety modified side chains at the 1-, 4-, 5-, 6-, and 7- positions led to the production of deoxyviolacein analogs (**7**, **10** or **11**). Reactions added with 5-fluoro-tryptophan (**1j**) and 7-fluoro-tryptophan (**1o**) resulted in the highest production of deoxyviolacein analogs (**7**), albeit less than a fifth that of native deoxyviolacein (**7a**).

The α-H of L-tryptophan is crucial for the violacein pathway. Consequently, analogs with a methyl substituent at the α-carbon (**1b** and **1c)** resulted in no detectable downstream metabolites. This is consistent with the first step of the proposed pathway, during which IPA **2** production requires a double bond formation at the α-carbon. Thus, inserting methyl group at this position blocks the formation of IPA **2**.

Tryptophan analogs **1q** to **1z** with substitutions at the 5- or 7- position of l-tryptophan were used to probe the substrate scope of the violacein pathway (Figure 3.6). Most 7-substituted tryptophans resulted in CPA byproduct formation (**4** and **9**). The production of CPA

heterodimer (**4v**) from reaction added with 7-chlorotryptophan (**1v**) is at least three times more than the amount of CPA(**4a**) produced from reaction added with tryptophan, in concordance with previous findings.[104] Deoxyviolacein analogs (**7**, **10**, or **11**) were detected in reactions added with 7-methyl-tryptophan (**1h**), 7-fluoro-tryptophan (**1o**), and 7-chloro-tryptophan (**1v).** TX-TL reactions added with tryptophan analogs **1q** to **1z** resulted in no production of homodimeric violacein analogs (**8**), although heterodimeric violacein analogs were detected from reactions added with 7-methyl-tryptophan (**1h**), 7-fluoro-tryptophan (**1o**), and 7-chloro-tryptophan (**1v**).

Most 5-substituted tryptophan analogs resulted in measurable downstream metabolites (shown in Figure 3.7). Although 5-substitution precludes the formation of homodimeric violacein analogs (**8**), the substrates 5-fluoro-tryptophan (**1j**), 5-chloro-tryptophan (**1q**), 5-bromo-tryptophan (**1r**), 5-methyl-trytophan (**1d**), and 5-cyano-tryptophan (**1t**) did result in heterodimeric violacein analogs **13**. Since the byproducts (CPA and deoxyviolacein) produced from reactions added with 5-substituted tryptophan analogs (**1d**, **1j**, and **1q** to **1u**) were predominantly heterodimeric as well, we hypothesize that this is due to the substrate preference of vioB to couple tryptophan analogs with native tryptophan, rather than another equivalent of tryptophan analog. The spontaneous oxidation of 5-hydroxy-l-tryptophan (**1g**) and 5-amino-l-tryptophan (**1n**) led to no measurable downstream metabolites. The reaction with 5-formyl-tryptophan (**1s**) is particularly interesting because the formyl group on deoxyviolacein can serve as a handle for further reactions, such as with organometallic reagents.

The production of various violacein analogs in the cell-free TX-TL system by simply adding tryptophan analogs and linear DNA offers a new method to discover new drug candidates. Simultaneously expressing multiple pathway enzymes and producing compounds of interest provides a biochemical platform not only to investigate reaction mechanism, but also to expand the diversity of pathway metabolites. The violacein analogs produced in TX-TL can be extracted, purified, and tested for their potency against natural violacein products. A continuous-exchange system[65] will enable the scale-up production of that target compound, and the use of concentrated extract can also increase yield.[63]

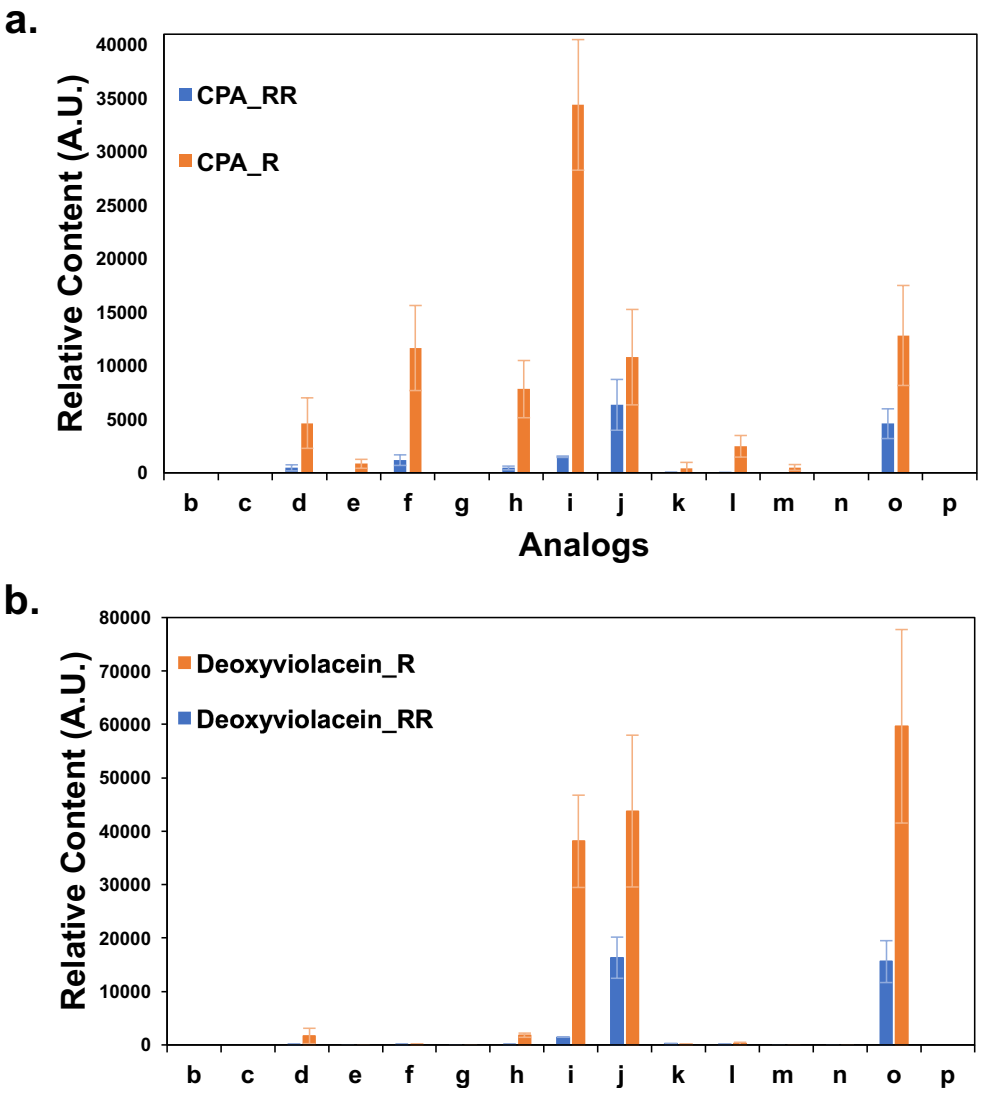


Figure 3.5: Analogs of the violacein pathway metabolites detected from TX-TL reactions. a. The signal of byproduct measured using LC-MS. b. The signal of deoxyviolacein analogs measured using LC-MS. CPA_R represents **9**, CPA_RR represent **4**, deoxyviolacein_R represent the combination of **10** and **11**, and deoxyviolacein_RR represent **7**.

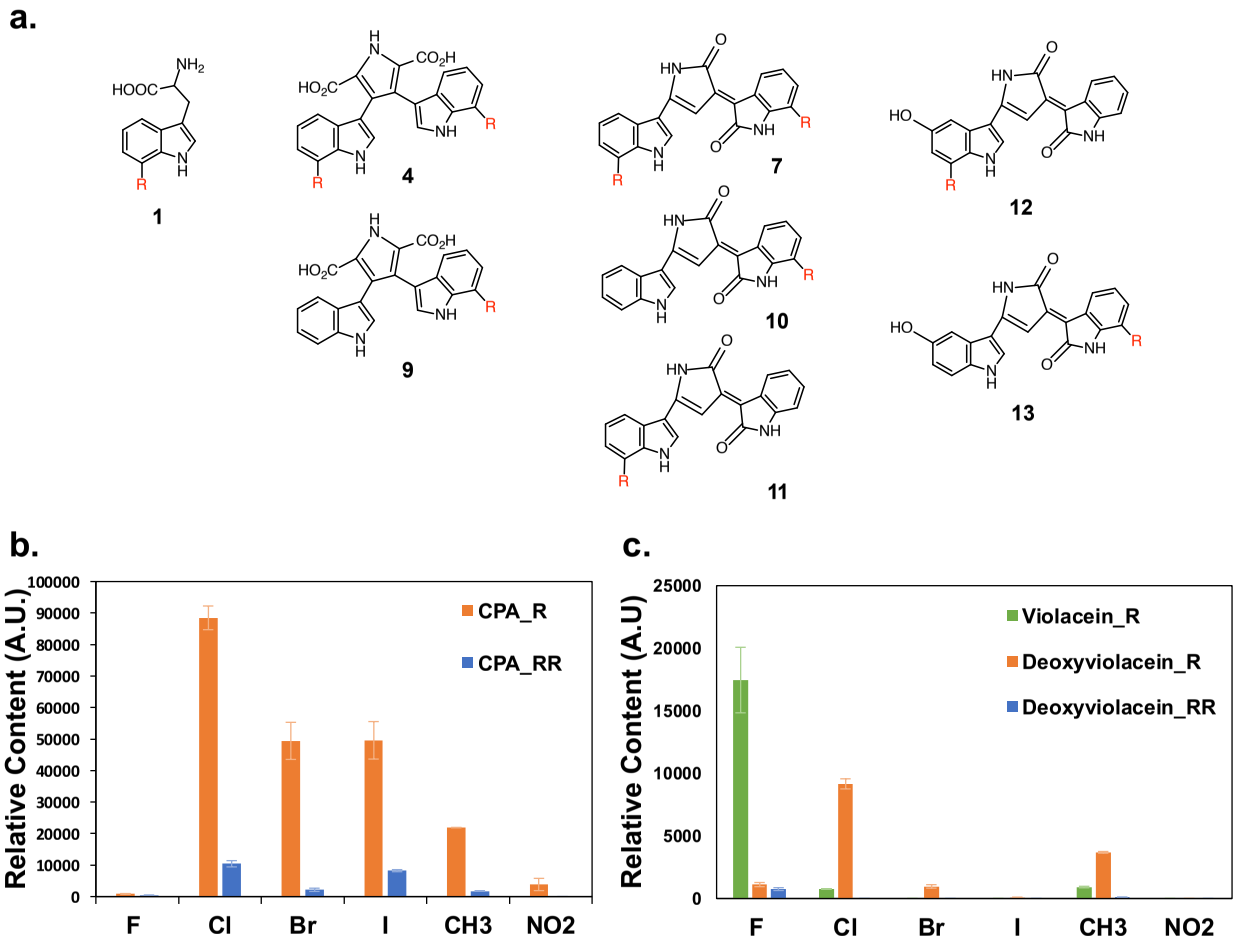


Figure 3.6: Analogs of the violacein pathway metabolites detected from TX-TL reactions added with 7-substituted tryptophans. a. Identified products from TX-TL reactions added with 7-substituted tryptophans. b. The signal of byproduct measured using LC-MS. c. The signal of violacein and deoxyviolacein analogs measured using LC-MS. CPA_R represents **9**, CPA_RR represent **4**, violacein R represent **13**, deoxyviolacein_R represent the combination of **10** and **11**, and deoxyviolacein_RR represent **7**.

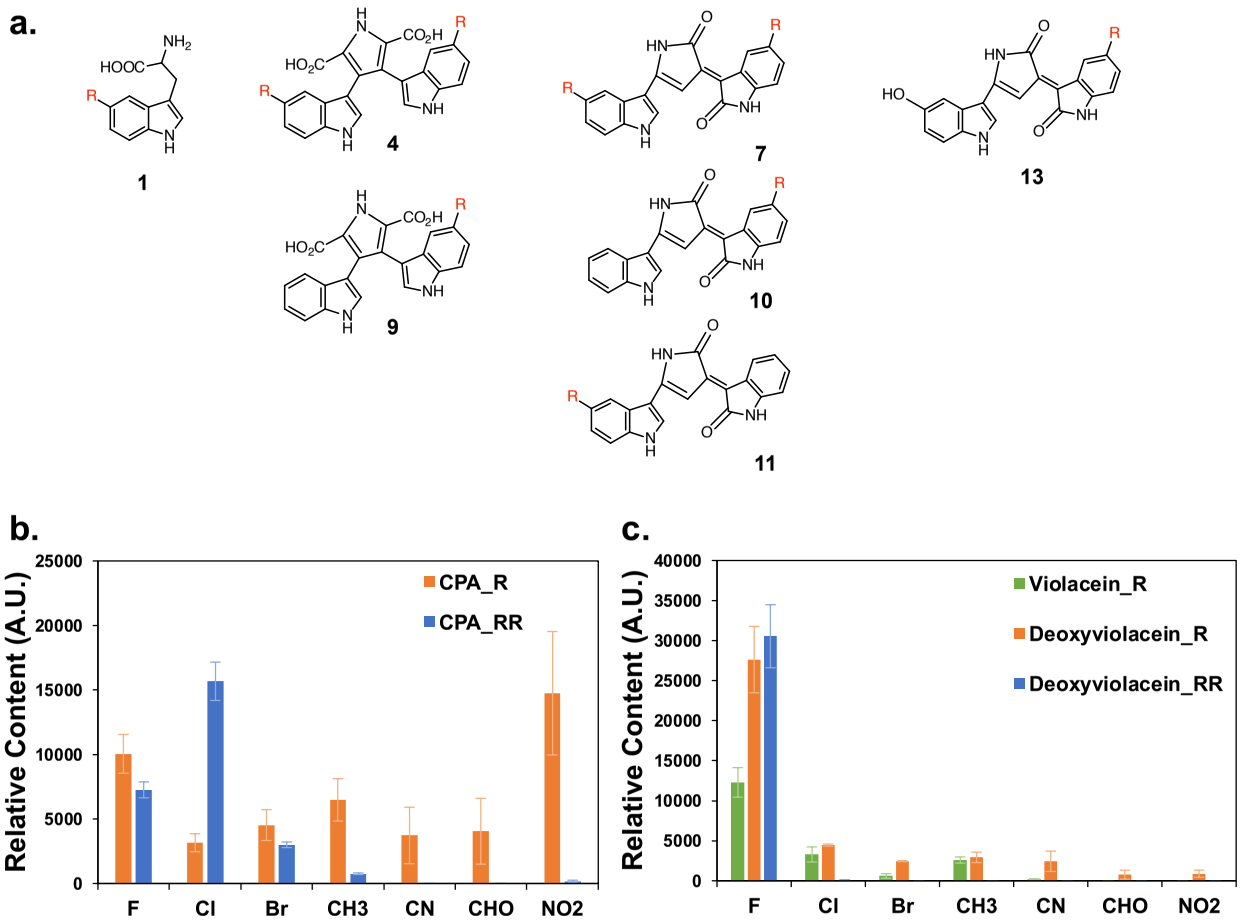


Figure 3.7: Analogs of the violacein pathway metabolites detected from TX-TL reactions added with 5-substituted-tryptophans. a. Identified products from TX-TL reactions added with 5-substituted tryptophans. b. The signal of byproduct measured using LC-MS. c. The signal of violacein and deoxyviolacein analogs measured using LC-MS. CPA_R represents **9**, CPA_RR represent **4**, violacein R represent **13**, deoxyviolacein_R represent **10**, and deoxyviolacein_RR represent **7**.

Protein Engineering

Even for substrates that led to violacein analogs, the CPA byproducts (**4** and **9**) were formed in significant quantities. Thus, engineering vioE to redirect metabolic flux towards target metabolites and away from byproduct formation will help further expand the utility of the TX-TL system. In some cases, reactions that produce only byproducts may be able to produce downstream metabolites with the help of an engineered vioE. In particular, engineering vioE for improved activity using a 7-fluoro-tryptophan (**1o**) can lead to the production of other 7-substituted violacein analogs.

Protein-Ligand Docking

Site saturated mutagenesis was used to generate a library of vioE variants, which were screened for improved activity with 7-fluoro-tryptophan (**1o**). In this work, the engineering of vioE was guided by protein-ligand docking using AutoDock Vina (Figure 3.8).[105] Of the five active-site residues,[106] Y17 was chosen for site-saturated mutagenesis because it resides close to intermediate **3**, but is not implicated in the proton transfer events involved in the 1,2-indole shift.

Results from protein-ligand docking also provide evidence to support the 1,2-indole shift mechanism. The positions of **3a** and **3o** both indicate that the nitrogen on one of the indoles is close to the catalytic residues and the nitrogen on the respective α-carbon is close to a water molecule. This suggests the catalytic residues facilitate the deprotonation step, while the water molecule facilitates the protonation step. The differences between a and b from

Figure 3.8 are the position of the carboxyl and the amine on the α-carbon of the other indole. The overlap of **3a** and **3o** in the vioE binding pocket is shown in Figure S3.7. The distances between the substrates and the residues of interests are shown in shown in Table S3.1. Since the crystal structure is previously obtained using PEG, the distance calculated may deviate from the actual values. However, these values shed light on the catalysis happening at the binding pocket of vioE and can guide the engineering of vioE.

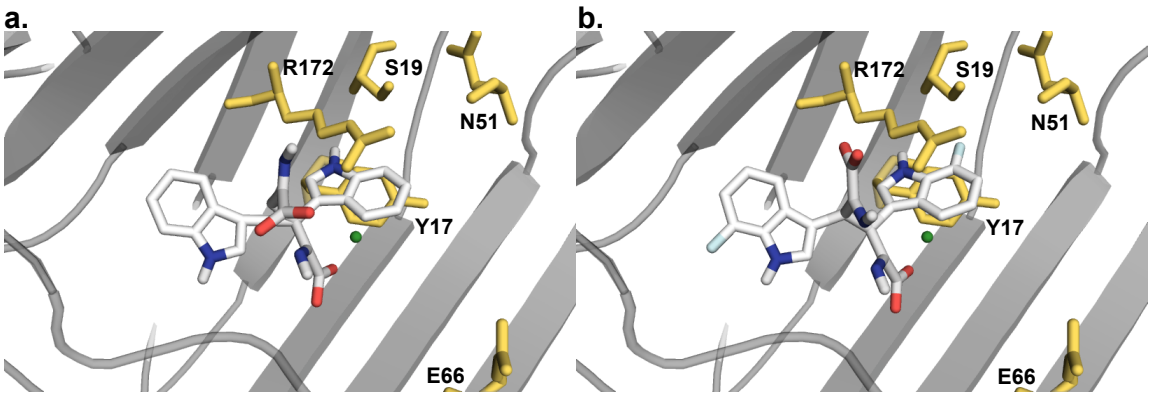


Figure 3.8: Results from protein-ligand docking--Interaction of intermediate **3a** with vioE (a) and intermediate **3o** with vioE (b) predicted using AutoDock Vina. Yellow—active sites, green—water molecule, blue—nitrogen, red—oxygen, turquoise—fluorine, and grey—vioE backbone.

Platform Validation

The cell-free TX-TL system delivers reproducible results and can become a reliable platform for protein engineering. TX-TL was evaluated as a screening method for engineering vioE for improved production of deoxyviolacein analogs. This work follows the procedures for screening platform validation detailed in previous studies.[107] Figure 3.9 shows the production of deoxyviolacein analog (**7o**) versus 34 wild-type vioE clones. Table 1 shows the rate of

false positives expected to find for screening libraries with different target thresholds and variability in the measurements (CV values). The false positive rates are calculated using a Gaussian distribution, and the fraction of false positives is calculated as the probability of finding a value of $\chi \geq \chi_0$ (see Supplementary Information for more details).

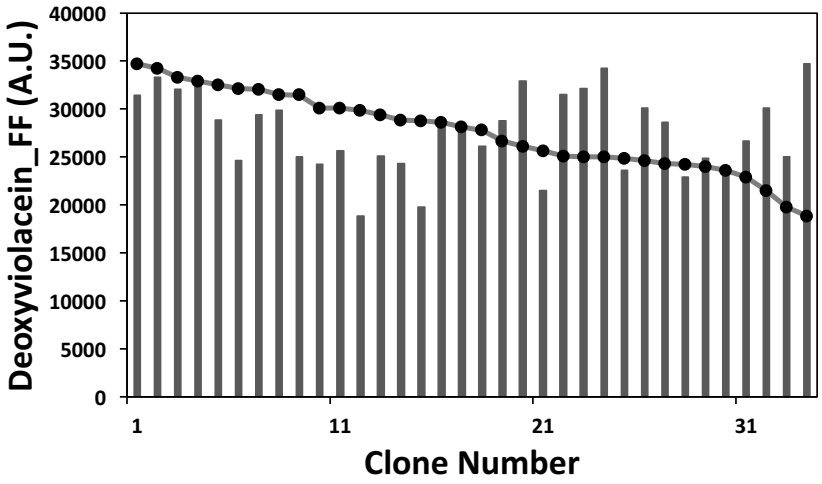


Figure 3.9: Evaluating TX-TL as a screening platform with wild-type samples. Enzyme activity is plotted versus reaction tubes and in descending order (dots). The CV for this screen is 15%.

Table 3.1: Rate of expected false positives for variants exhibiting higher than 1.5 and 2.0 times the mean μ.

Fitness target	1.5 μ	2.0 μ
CV	15%	15%
Fraction of false positives	4.3×10^{-4}	1.3×10^{-11}

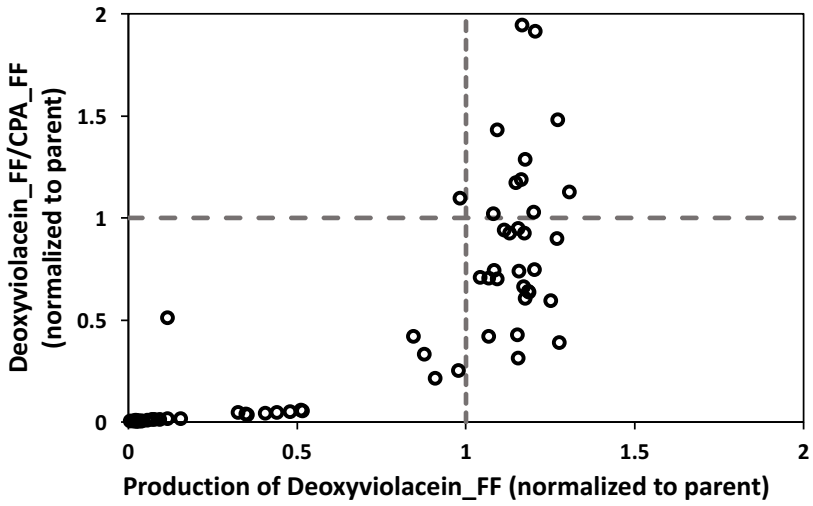


Figure 3.10: Analysis of a mutant library: Relationship between the metabolic flux split between CPA analog and deoxyviolacein analog and the production of deoxyviolacein analog. Deoxyviolacein_FF represents **7o**, and CPA_FF represent CPA **4o**.

Results from site-directed mutagenesis of Y17 are shown in Figure 3.10. The metabolic flux split between byproduct CPA analog and deoxyviolacein analog is characterized using the ratio of deoxyviolacein analog **7o** over CPA analog **4o**. The top two mutants that have the highest ratio of **7o** over **4o** are both Y17H. They direct the metabolic flux from the byproduct CPA analog towards the production of deoxyviolacein analog. The ratio of deoxyviolacein **7o** over CPA **4o** is twice the wild-type value. Although the production of deoxyviolacein analog shows no significant improvement, we hypothesize that the metabolic flux has been directed to another downstream byproduct deoxychromoviridans. Based on the previously proposed mechanism (shown in Figure S3.11), we hypothesize that the lack of oxygen control in the cell-free TX-TL system inhibits the improvement in the production of deoxyviolacein analog. On the other hand, the bottom two mutants are Y17L. Y17 was previously identified to be deleterious during alanine scanning,[106] and

leucine shares a similar structure with alanine. Nevertheless, the ability to create a library of protein mutants for improved catalytic activity demonstrates that TX-TL is a viable platform for protein engineering.

CONCLUSIONS AND FUTURE DIRECTIONS

In this work, we have studied the substrate scope of the violacein pathway and found pathway enzymes capable of catalyzing fortuitous reactions. We find the violacein pathway enzymes accept substrates that are structurally similar to l-tryptophan with decorated indole side chains. Our experimental results are consistent with the pathway reaction mechanism that was previously proposed based on *in vitro* experiments using purified enzymes. Our results from molecular docking is also consistent with the pathway reaction mechanism. More tryptophan analogs are needed to validate the pathway reaction mechanism further, but this work has shown that TX-TL can serve as a biochemical platform for characterizing and engineering biosynthetic pathways. We have evaluated and verified TX-TL as a screening method for enzyme activity, which suggests TX-TL as a reliable alternative platform for protein engineering. We have observed a diverse portfolio of products in the cell-free TX-TL system, which can be applied to further medicinal research. We have yet to mix and match more than two different tryptophan analogs in TX-TL reactions, which may further expand the substrate scope of the violacein pathway.

MATERIALS AND METHODS

TX-TL Reactions Extract Making

Collected cell pellets are homogenized using Microfluidizer Processor and extracted according to the method described by Kwon *et al.*[86] with the post-homogenization incubation period extended to 80 min instead of 60 min. Buffer preparation is done according to protocol developed in previous studies[33] with a supplement of 15 mM maltose. The preparation results in extract with conditions: 8.9−9.9 mg/mL protein, 4.5−10.5 mM Mg-glutamate, 100—140 mM K-glutamate, 0.33−3.33 mM DTT, 1.5 mM each amino acid except leucine, 1.25 mM leucine, 50 mM HEPES, 1.5 mM ATP and GTP, 0.9 mM CTP and UTP, 0.2 mg/mL tRNA, 0.26 mM CoA, 0.33 mM NAD+, 0.75 mM cAMP, 0.068 mM folinic acid, 1 mM spermidine, 30 mM 3-PGA, 0.15 mM NADPH, and 1 mM acetyl-coA. For experiments requiring no additional tryptophan, water is added instead of tryptophan in the amino acid mix. When needed, linear DNA or plasmid DNA is added to a mix of extract and buffer. TX-TL reactions are conducted in Microcentrifuge tubes and kept at 29°C. BioTeK Synergy H1 microplate reader is used to collect kinetic data for a fluorescent protein, if necessary.

Protein Docking

We apply the crystal structure of vioE (PDB ID: 3BMZ)[106] to Autodock Vina[105] to investigate the potential interactions between substrates and vioE. The substrate used for crystallography is removed from the crystal structure, and a ".pdbqt" file is generated for further processing with Autodock Vina. Substrates used for protein docking are **3a**, **3d**, and **3h**, and their

structures were created using PRODRG.[108] We follow docking procedures detailed in previous protocol,[109] with the docking area of interests is set to cover the entire protein. The final images are generated using PyMOL by opening the generated files in the same session.

Library Creation and Screening

Site-saturated mutagenesis is carried out using the 22-c trick.[110] For a given site targeted for mutagenesis, three primers are designed containing codons NDT, VHG, and TGG respectively. The three primers are mixed in a ratio of 12:9:1. The QuikChange II Site-Directed Mutagenesis Kit is used for library creation. PCR is carried out using PfuUltra HF DNA polymerase. A plasmid that contained the parent gene in the pBEST vector as a template.[102] The sample is digested with *Dpn*I and transformed in JM109. At least 66 colonies are picked from each site saturation library to proceed for colony PCR. The PCR products are directly added into TX-TL reactions for simultaneous enzyme expression and metabolite production. Linear DNA encoding other pathway enzymes are added for screening the production of deoxyviolacein analogs in TX-TL. Protein gamS is added into TX-TL reactions to prevent linear DNA degradation.[60]

Linear DNA

PCR products are amplified using Q5® High-Fidelity 2X Master Mix (New England BioLabs). PCR products undergo an additional PCR purification step using DNA Clean & Concentrator™-5-Capped Columns (Zymo Research), which removes excess salt detrimental to TX-TL and are eluted and stored in water at −20°C for long-term storage.

LC-MS

UPLC-MS is used to analyze TX-TL samples. TX-TL samples were first diluted with 1:19 volume ratio with methanol to remove proteins and other big molecules before analysis. Xevo G2-S QTof Quadrupole Time-of-Flight Mass Spectrometry, interfaced with UPLC, utilizing electrospray ionization (ESI) and MRM based acquisition methods are used. Compounds are detected using negative ionization mode. Waters XBridge BEH C18 2.1x100mm (particle size 2.5 um) Column XP was used. The column temperature is maintained at 45°C with a flow rate of 0.4 mL/min. Injection volume is 2.5 ul. Eluents include water with 0.1% formic acid (A) and acetonitrile (B). A fast 5 min 20-80% acetonitrile gradient is used, resulting in a 7-minute long LC-MS method.

ACKNOWLEDGEMENTS

We thank Michael E. Lee and the Dueber Lab from the University of California, Berkeley for providing us plasmid with the violacein pathway. We thank Dr. Nathan Dalleska and the Environmental Analysis Center for the support and assistance using UPLC/MS for data collection. We thank Sheel Dodani, Shan Huang, Andrew Buller, Rusty Lewis, and Anders Knight for helpful discussion. Y.Y.W was supported by NIH/NRSA Training Grant 5 T32 GM07616 and the Gordon and Betty Moore Foundation Grant GBMF2809 to the Caltech Programmable Molecular Technology Initiative.

CONFLICT OF INTERESTS

R.M.M is a co-founder and board member of Synvitrobio, a for-profit company developing cell-free systems.

SUPPLEMENTARY INFORMATION

Tryptophan Notation

Tryptophan analogs with various substitutions on L-tryptophan are used in this study. Figure S3.1 shows the numbering of positions on the indole moiety used for communication of this work. The numbering starts with the nitrogen position and goes counterclockwise around the indole ring.

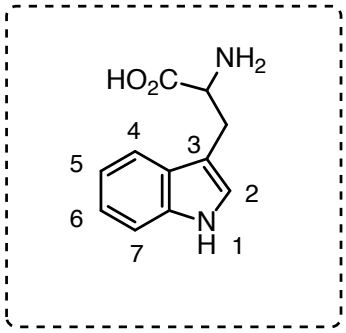


Figure S3.1: Number of the positions on the indole moiety of tryptophan.

Enzyme Expression in TX-TL

Pathway enzyme expression in TX-TL is verified using Western blotting (shown in Figure S3.2). From left to right is a protein ladder, negative control, vioA, vioB, vioC, vioD, and vioE. The negative control is a TX-TL reaction where water is added instead of DNA. The following samples are TX-TL reactions added with linear DNA encoding corresponding enzymes. The Western blot protocol used is previously optimized for proteins around 50 kDa, which explains why enzymes of lower or higher size range do not show well on the Western blot. The constructs used in this work come from previous work.[102]

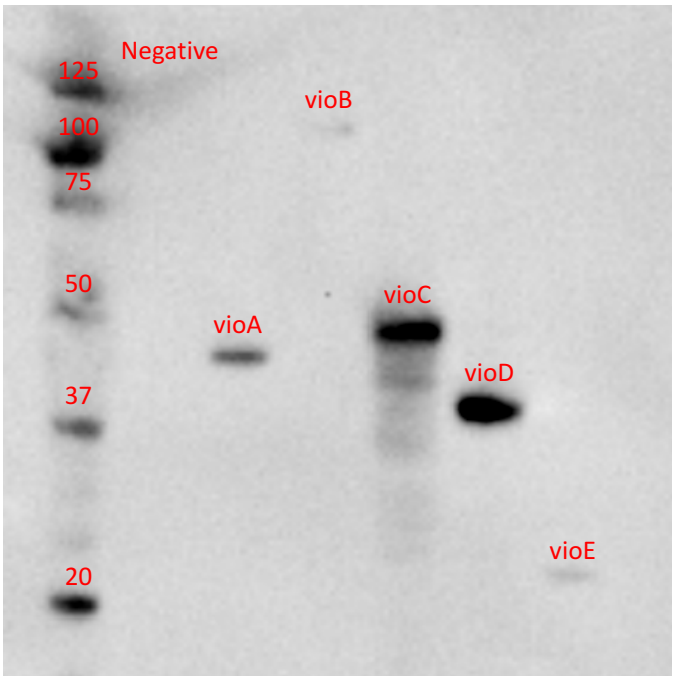


Figure S3.2: Enzyme expression of the violacein pathway.

Docking Results of vioE

The docking results of vioE with substrates **3a** and **3h** are shown in Figure S3.3 and Figure S3.4. AutoDock Vina predicts two different positions for each substrate. The first position from each prediction is the one at which the predicted binding energy is minimized. The first position predicted for intermediate **3a** overlaps with the first position predicted for intermediate **3h** very closely (shown in Figure S3.5a). The second position predicted for intermediate **3a** deviates from the second position predicted for intermediate **3h,** which may suggest that the first position predicts the catalytic interactions between substrates and vioE. The two positions might be very different, but they both capture the position at which the active sites interact with half of the intermediate, where the 1, 2-indole shift occurs. The

distances of substrates **3a** and **3h** to nearby active sites are measured in PyMOL (shown in Table S3.1). For the following discussion, the hydrogen attached to the nitrogen that undergoes protonation is considered H1, and the hydrogen that undergoes deportation is considered H2. The distance between H1 and the water molecule is the same for the first predicted positions. For the four different positions predicted, the distance of H2 to active sites are very similar. For all four positions, the distance between H2 and Y17 is the shortest, and the distance calculated between H1 and the water molecule is the second shortest.

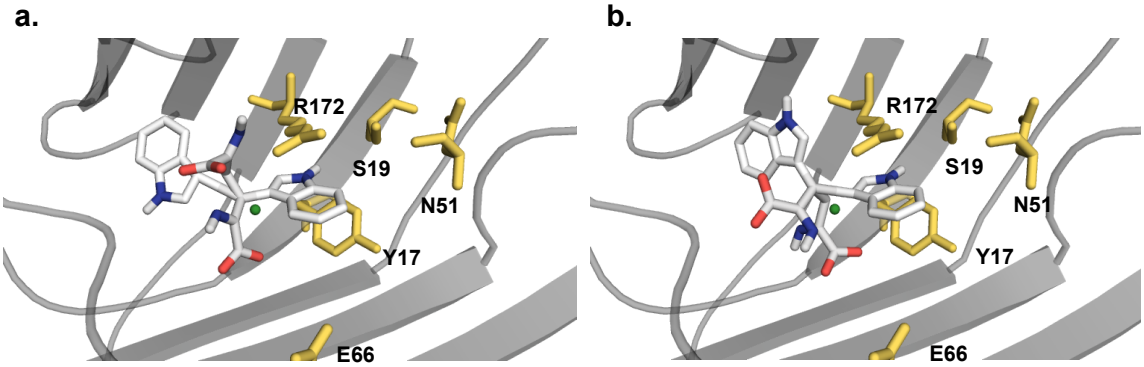


Figure S3.3: The interaction of intermediate **3a** and vioE predicted with AutoDock Vina. Yellow—active sites, green—water molecule, blue—nitrogen, red—oxygen, and grey—vioE backbone.

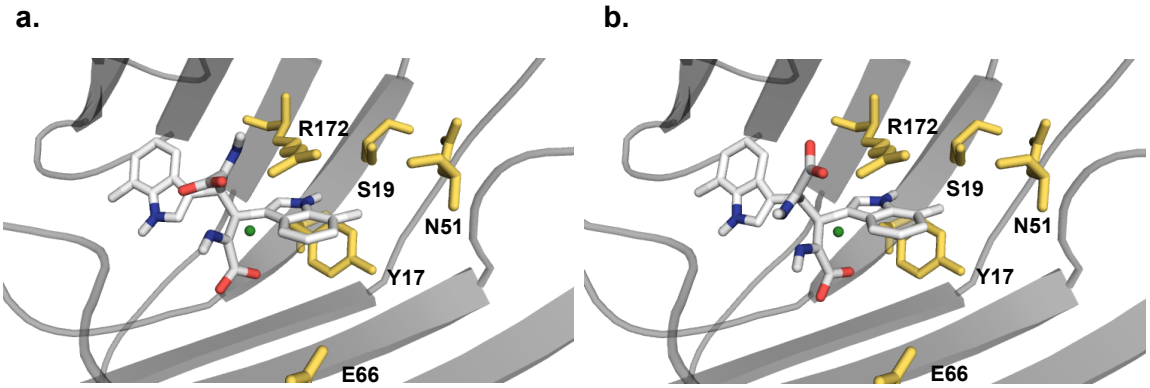


Figure S3.4: The interaction of intermediate **3h** and vioE predicted with AutoDock Vina. Yellow—active sites, green—water molecule, blue—nitrogen, red—oxygen, and grey—vioE backbone.

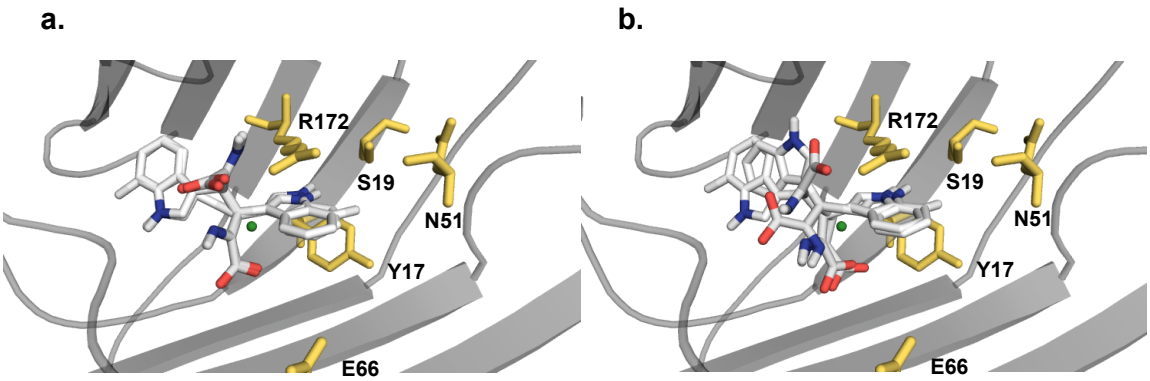


Figure S3.5: The interaction of intermediate **3a** and vioE overlapped with the interaction of intermediate **3h** and vioE predicted with AutoDock Vina. Yellow—active sites, green—water molecule, blue—nitrogen, red—oxygen, and grey—vioE backbone.

Docking results for all three intermediates capture the position at which the active sites interact with half of the intermediate, where the 1,2-indole shift occurs. The mechanism previously proposed[101] is redrawn and shown in Figure S3.11.

The overlapping docking results of vioE with substrates **3a** and **3o** are shown in Figure S3.7. The interaction of **3o** with vioE are shown in Figure S3.6. Again, the distance between H1 and the water molecule is the same for the first predicted positions (shown in Table S3.1). For the four different positions predicted, the distance of H2 to active sites are also very similar. The predicted positions of **3o** and **3h** are different. Based on the similarity to the position of **3a**, **3h** may have a more similar interaction with vioE than **3o** may have. However, the low production level of deoxyviolacein analogs **7h** has been detected. In contrast, the production level of deoxyviolacein analog **7o** is about ten times higher than the production level of deoxyviolacein analog **7h**.

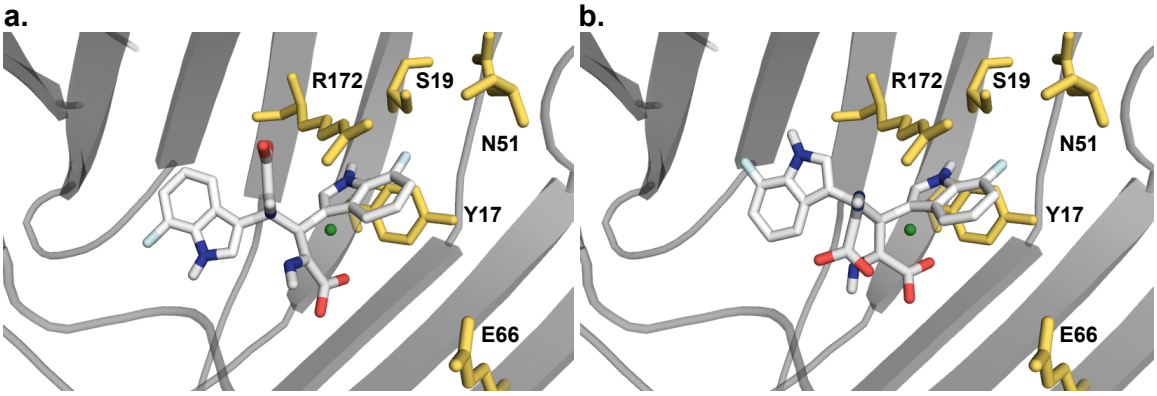


Figure S3.6: The interaction of intermediate **3o** and vioE predicted with AutoDock Vina. Yellow—active sites, Green—water molecule, blue—nitrogen, red—oxygen, turquoise—fluorine, and grey—vioE backbone.

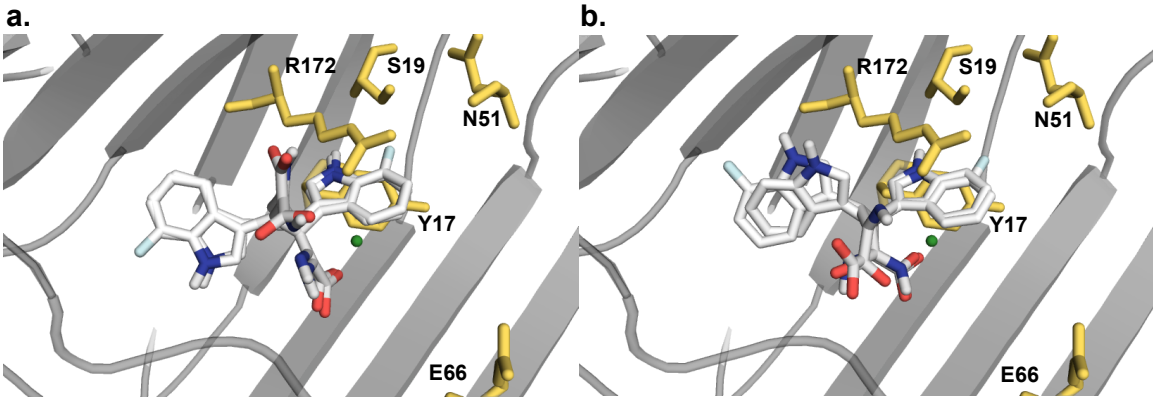


Figure S3.7: The interaction of intermediate **3a** and vioE overlapped with the interaction of intermediate **3o** and vioE predicted with AutoDock Vina. Yellow—active sites, Green—water molecule, blue—nitrogen, red—oxygen, turquoise—fluorine, and grey—vioE backbone.

Table S3.1: Distance between substrates and molecules of interests.

	Distance from Nitrogen to Nearby Molecules of Interests (Å)				
Substrate	Water	Arginine	Tyrosine	Asparagine	Serine
3a	2.8	2.8	2.4	7.4	4.0
	2.2	2.9	2.2	7.1	3.7
3h	2.8	3.1	2.2	7.0	3.8
	2.8	2.9	2.5	7.6	4.8
3o	2.8	2.8	2.5	7.6	4.3
	5.2	3.2	2.2	6.8	3.7

TX-TL as a Platform for Protein Engineering

Figure S3.8 shows the relationship between the production of deoxyviolacein analog **7o** and DNA purity. DNA purity, defined by the ratio of absorbance at 260 over the absorbance at 230, need to be greater than 1.6 to minimize the possibility of screening a false positive. Previous experiments show that DNA purity lower than 1.6 results with no production of deoxyviolacein analog.

TX-TL reactions with the omission of additional tryptophan significantly eliminate the competition with the native substrate and reduce the amount of heterodimer formation (shown in Figure S3.10, R = 7-fluoro). The extract used for this work has not been treated with dialysis, and a significant amount of tryptophan (4 mM) is present. The additional tryptophan from the buffer supplements for enzyme expression and the omission of tryptophan may impact enzyme expression. However, result from control experiment with GFP shows no reduction in protein expression (shown in Figure S3.9). Nevertheless, the metabolite production level of TX-TL reaction with no additional tryptophan is lower than

the metabolite production level of TX-TL reaction with additional tryptophan. We hypothesize that fluorotryptohans are incorporated in protein synthesis. Adding tryptophan analogs after the first two hour of TX-TL reactions may eliminate the incorporation of non-canonical amino acids into protein synthesis, which may help maintaining similar metabolite production level in TX-TL reactions with and without additional tryptophan.

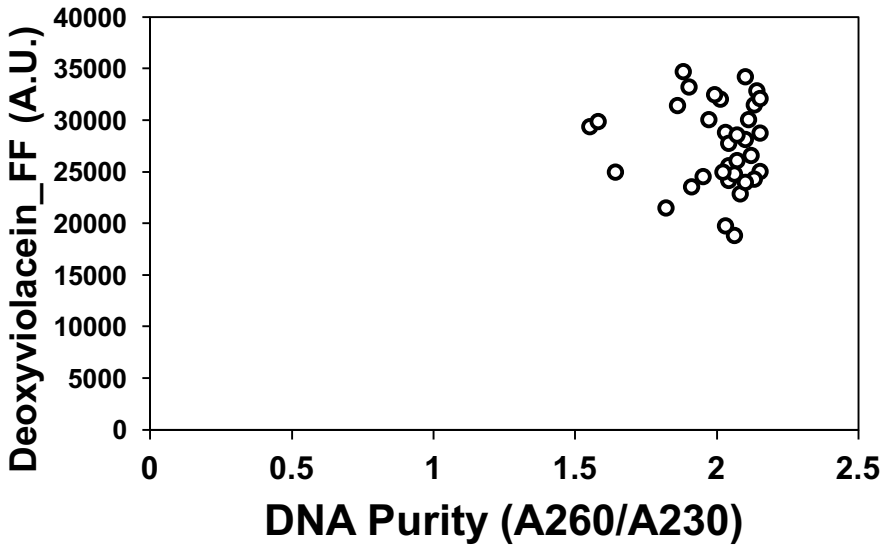


Figure S3.8: The production of deoxyviolacein analog **7o** against DNA purity.

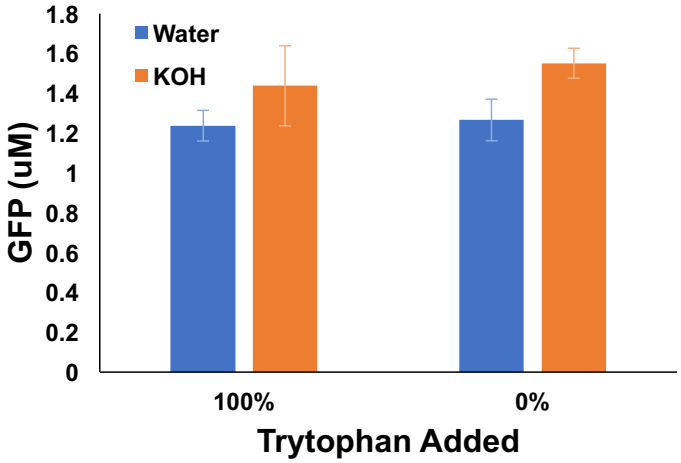


Figure S3.9: GFP production in TX-TL reactions with and without additional tryptophan. Blue bars: reactions added with water. Orange bars: reactions added with tryptophan analogs dissolved in 5mM of KOH.

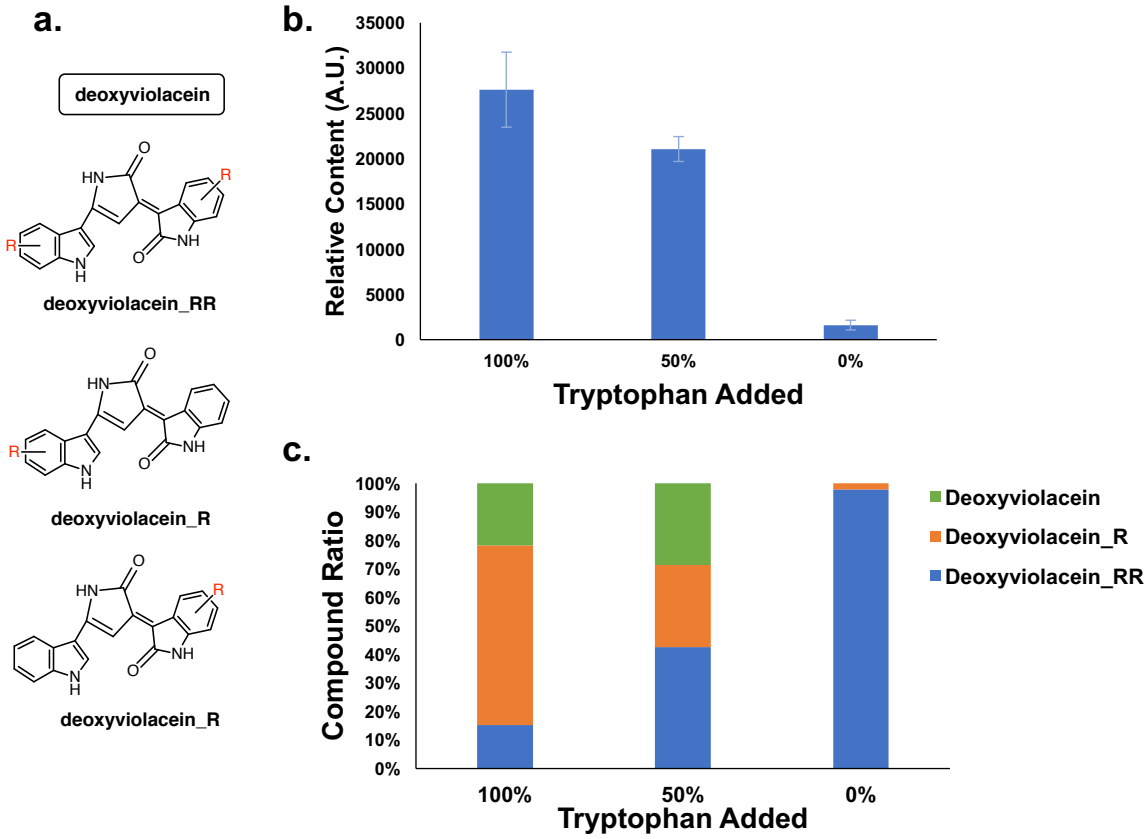


Figure S3.10: Deoxyviolacein analog production in TX-TL added with and without additional tryptophans. a. The possible deoxyviolacein analogs. b. The relative amount of the homodimer of deoxyviolacein analog produced with and without additional tryptophan. c. The ratio of various deoxyviolacein analog produced with and without additional tryptophan.

Validating TX-TL as Screening Platform

TX-TL is evaluated as a screening platform in this work. The false positive rates are calculated using a Gaussian distribution, and the fraction of false positives is calculated as the probability of finding a value of $\chi \geq \chi_0$. The probability function of a Gaussian distribution is:

$$F(x) = \frac{1}{2\pi}\int_0^x \frac{-(x-\mu)^2}{2\sigma^2} dx \qquad \text{(S3.E.1)}$$

where μ represents the mean of the distribution, σ represents the standard deviation, and *F(x)* represents the probability of a random variable with values less than or equal to *x*. The rate of false positives is calculated using

$$rate = 1 - F\left(\frac{\chi_0 - \mu}{\sigma}\right) \quad \text{(S3.E.2)}$$

where $F(\frac{\chi_0 - \mu}{\sigma})$ is calculated using S3.E.1. Values in Table 3.1 are calculated using Microsoft Excel function NORM.DIST.

Compound Detection

The $[M-H]^-$ formula of a given compound is shown in Table S3.2. The required m/z under ESI for the compounds are shown in Table S3.4. The observed m/z for detected compounds are shown in Table S3.5, and the element composition is validated by high-resolution mass spectrometry (HRMS) and single mass analysis. Figure S3.12 and Figure S3.13 show the analysis of deoxyviolacein analog **7o** produced in reactions added with 7-fluoro-tryptophan **1o** and deoxyviolacein analog produced in reactions added with 5-formyl-tryptophan **1o** (b). Table S3.3 shows the retention time of each compound. It is difficult to separate the two different possible heterodimers of deoxyviolacein analogs, but we have observed different retention times for the two possible heterodimers in reactions added with 7-chloro-tryptophan (**1v**). Compound IPA (**2**) and intermediate (**3**) are too unstable to be detected. Intermediate prodeoxyviolacein (**5**) is omitted for the tables below. "n/a" represents either the compound is not possible based on known reaction mechanism or not observed during this study.

		Numbering	1	1	4	6	7	8
#	Alphabet	Position	Tryptophan	Trytophan	CPA	Proviolacein	Deoxyviolacein	Violacein
0	a	0	L-trytophan	$C_{11}H_{11}N_2O_2$	$C_{22}H_{14}N_3O_4$	$C_{20}H_{12}N_3O_2$	$C_{20}H_{12}N_3O_2$	$C_{20}H_{12}N_3O_3$
1	b	3	α-methyl-L-tryptophan	$C_{12}H_{13}N_2O_2$	n/a	n/a	n/a	n/a
2	c	3	α-methyl-D-tryptophan	$C_{12}H_{13}N_2O_2$	n/a	n/a	n/a	n/a
3	d	5	5-methyl-DL-tryptophan	$C_{12}H_{13}N_2O_2$	$C_{24}H_{18}N_3O_4$	$C_{22}H_{16}N_3O_2$	$C_{22}H_{16}N_3O_2$	n/a
4	e	Naphthalene	3-(2-naphthyl)-L-alanine	$C_{13}H_{12}NO_2$	n/a	n/a	n/a	n/a
5	f	1	1-methyl-L-tryptophan	$C_{12}H_{13}N_2O_2$	$C_{24}H_{18}N_3O_4$	$C_{22}H_{16}N_3O_2$	$C_{22}H_{16}N_3O_2$	n/a
6	g	5	5-hydroxy-L-tryptophan	$C_{11}H_{11}N_2O_3$	n/a	n/a	n/a	n/a
7	h	7	7-methyl-DL-tryptophan	$C_{12}H_{13}N_2O_2$	$C_{24}H_{18}N_3O_4$	$C_{22}H_{16}N_3O_2$	$C_{22}H_{16}N_3O_2$	n/a
8	i	6	6-fluoro-DL-tryptophan	$C_{11}H_{10}FN_2O_2$	$C_{22}H_{12}F_2N_3O_4$	n/a	$C_{20}H_{10}F_2N_3O_2$	n/a
9	j	5	5-fluror-DL-tryptophan	$C_{11}H_{10}FN_2O_2$	$C_{22}H_{12}F_2N_3O_4$	n/a	$C_{20}H_{10}F_2N_3O_2$	n/a
10	k	4	4-methyl-DL-tryptophan	$C_{12}H_{13}N_2O_2$	$C_{24}H_{18}N_3O_4$	$C_{22}H_{16}N_3O_2$	$C_{22}H_{16}N_3O_2$	n/a
11	l	4	4-fluoro-DL-tryptophan	$C_{11}H_{10}FN_2O_2$	$C_{22}H_{12}F_2N_3O_4$	n/a	n/a	n/a
12	m	1	L-3-benzothienylalanine	$C_{11}H_{10}NO_2S$	n/a	n/a	n/a	n/a
13	n	5	5-amino-L-tryptophan	n/a	n/a	n/a	n/a	n/a
14	o	7	7-fluoro-DL-tryptophan	$C_{11}H_{10}FN_2O_2$	$C_{22}H_{12}F_2N_3O_4$	n/a	$C_{20}H_{10}F_2N_3O_2$	n/a
15	p	3	L-Tryptophanamide	n/a	n/a	n/a	n/a	n/a
16	q	5	5-chloro-L-tryptophan	$C_{11}H_{10}ClN_2O_2$	$C_{22}H_{12}Cl_2N_3O_4$	n/a	$C_{20}H_{10}Cl_2N_3O_2$	n/a
17	r	5	5-bromo-L-tryptophan	$C_{11}H_{10}BrN_2O_2$	$C_{22}H_{12}Br_2N_3O_4$	n/a	n/a	n/a
18	s	5	5-formyl-L-tryptophan	$C_{12}H_{11}N_2O_3$	n/a	n/a	n/a	n/a
19	t	5	5-cyano-L-tryptophan	$C_{12}H_{10}N_3O_2$	n/a	n/a	n/a	n/a
20	u	5	5-nitro-L-tryptophan	$C_{11}H_{10}N_3O_4$	$C_{22}H_{12}N_5O_8$	n/a	n/a	n/a
21	v	7	7-chloro-L-tryptophan	$C_{11}H_{10}ClN_2O_2$	$C_{22}H_{12}Cl_2N_3O_4$	n/a	n/a	n/a
22	w	7	7-bromo-L-tryptophan	$C_{11}H_{10}BrN_2O_2$	$C_{22}H_{12}Br_2N_3O_4$	n/a	n/a	n/a
23	x	7	7-iodo-L-tryptophan	$C_{11}H_{10}IN_2O_2$	$C_{22}H_{12}I_2N_3O_4$	n/a	n/a	n/a
24	y	7	7-cyano-L-tryptophan	$C_{12}H_{10}N_3O_2$	n/a	n/a	n/a	n/a
25	z	7	7-nitro-L-tryptophan	$C_{11}H_{10}N_3O_4$	n/a	n/a	n/a	n/a

Table S3.2 Formula of compounds validated by HRMS.

		Numbering	9	10	11	12	13
#	Alphabet	position	CPA _hetero	Deoxyviolacein _hetero_L	Deoxyviolacein _hetero_R	Violacein _hetero_L	Violacein _hetero_R
0	a	0	n/a	n/a	n/a	n/a	n/a
1	b	3	n/a	n/a	n/a	n/a	n/a
2	c	3	n/a	n/a	n/a	n/a	n/a
3	d	5	$C_{23}H_{16}N_3O_4$	$C_{21}H_{14}N_3O_2$	$C_{21}H_{14}N_3O_2$	n/a	$C_{21}H_{14}N_3O_3$
4	e	Naphthalene	$C_{24}H_{15}N_2O_4$	n/a	n/a	n/a	n/a
5	f	1	$C_{23}H_{16}N_3O_4$	$C_{21}H_{14}N_3O_2$	$C_{21}H_{14}N_3O_2$	n/a	n/a
6	g	5	n/a	n/a	n/a	n/a	n/a
7	h	7	$C_{23}H_{16}N_3O_4$	$C_{21}H_{14}N_3O_2$	$C_{21}H_{14}N_3O_2$	$C_{21}H_{14}N_3O_3$	$C_{21}H_{14}N_3O_3$
8	i	6	$C_{22}H_{13}FN_3O_4$	$C_{20}H_{11}FN_3O_2$	$C_{20}H_{11}FN_3O_2$	n/a	n/a
9	j	5	$C_{22}H_{13}FN_3O_4$	$C_{20}H_{11}FN_3O_2$	$C_{20}H_{11}FN_3O_2$	n/a	$C_{20}H_{11}FN_3O_3$
10	k	4	$C_{23}H_{16}N_3O_4$	$C_{21}H_{14}N_3O_2$	$C_{21}H_{14}N_3O_2$	n/a	n/a
11	l	4	$C_{22}H_{13}FN_3O_4$	$C_{20}H_{11}FN_3O_2$	$C_{20}H_{11}FN_3O_2$	n/a	n/a
12	m	1	$C_{22}H_{13}N_2O_4S$	n/a	n/a	n/a	n/a
13	n	5	n/a	n/a	n/a	n/a	n/a
14	o	7	$C_{22}H_{13}FN_3O_4$	$C_{20}H_{11}FN_3O_2$	$C_{20}H_{11}FN_3O_2$	$C_{20}H_{11}FN_3O_3$	$C_{20}H_{11}FN_3O_3$
15	p	3	n/a	n/a	n/a	n/a	n/a
16	q	5	$C_{22}H_{13}ClN_3O_4$	$C_{20}H_{11}ClN_3O_2$	$C_{20}H_{11}ClN_3O_2$	n/a	$C_{20}H_{11}ClN_3O_3$
17	r	5	$C_{22}H_{13}BrN_3O_4$	$C_{20}H_{11}BrN_3O_2$	$C_{20}H_{11}BrN_3O_2$	n/a	$C_{20}H_{11}BrN_3O_3$
18	s	5	$C_{23}H_{14}N_3O_5$	$C_{21}H_{12}N_3O_3$	$C_{21}H_{12}N_3O_3$	n/a	n/a
19	t	5	$C_{23}H_{13}N_4O_4$	$C_{21}H_{11}N_4O_2$	$C_{21}H_{11}N_4O_2$	n/a	$C_{21}H_{11}N_4O_3$
20	u	5	$C_{22}H_{13}N_4O_6$	$C_{20}H_{11}N_4O_4$	$C_{20}H_{11}N_4O_4$	n/a	n/a
21	v	7	$C_{22}H_{13}ClN_3O_4$	$C_{20}H_{11}ClN_3O_2$	$C_{20}H_{11}ClN_3O_2$	$C_{20}H_{11}ClN_3O_2$	$C_{20}H_{11}ClN_3O_3$
22	w	7	$C_{22}H_{13}BrN_3O_4$	$C_{20}H_{11}BrN_3O_2$	$C_{20}H_{11}BrN_3O_2$	n/a	n/a
23	x	7	$C_{22}H_{13}IN_3O_4$	$C_{20}H_{11}IN_3O_2$	$C_{20}H_{11}IN_3O_2$	n/a	n/a
24	y	7	$C_{23}H_{13}N_4O_4$	n/a	n/a	n/a	n/a
25	z	7	$C_{22}H_{13}N_4O_6$	n/a	n/a	n/a	n/a

Table S3.2 Formula of compounds validated by HRMS (continued).

		Numbering	1	4	6	7	8
#	alphabet	position	tryptophan	CPA	Proviolacein	Deoxyviolacein	Violacein
0	a	0	0.80	2.28	1.70	2.89	2.31
1	b	3	0.84	n/a	n/a	n/a	n/a
2	c	3	0.84	n/a	n/a	n/a	n/a
3	d	5	0.98	2.81	2.55	3.05	n/a
4	e	Naphthalene	1.17	n/a	n/a	n/a	n/a
5	f	1	1.01	3.08	2.65	3.04	n/a
6	g	5	0.74	n/a	n/a	n/a	n/a
7	h	7	0.97	2.81	2.55	3.05	n/a
8	i	6	0.88	2.56	n/a	3.34	n/a
9	j	5	0.87	2.42	n/a	3.22	n/a
10	k	4	0.97	n/a	2.46	3.04	n/a
11	l	4	0.86	n/a	n/a	n/a	n/a
12	m	1	1.05	n/a	n/a	n/a	n/a
13	n	5	n/a	n/a	n/a	n/a	n/a
14	o	7	0.87	2.42	n/a	3.23	n/a
15	p	3	n/a	n/a	n/a	n/a	n/a
16	q	5	1.10	2.89	n/a	3.87	n/a
17	r	5	1.17	3.04	n/a	n/a	n/a
18	s	5	0.76	n/a	n/a	n/a	n/a
19	t	5	0.80	n/a	n/a	n/a	n/a
20	u	5	0.87	2.17	n/a	n/a	n/a
21	v	7	1.06	3.05	n/a	n/a	n/a
22	w	7	1.13	3.21	n/a	n/a	n/a
23	x	7	1.25	3.45	n/a	n/a	n/a
24	y	7	0.84	n/a	n/a	n/a	n/a
25	z	7	0.88	n/a	n/a	n/a	n/a

Table S3.3: Retention time of detected compounds (in minutes).

		Numbering	9	10	11	12	13
#	alpha	position	CPA _hetero	Deoxyviolacein _hetero_L	Deoxyviolacein _hetero_R	Violacein _hetero_L	Violacein _hetero_R
0	a	0	n/a	n/a	n/a	n/a	n/a
1	b	3	n/a	n/a	n/a	n/a	n/a
2	c	3	n/a	n/a	n/a	n/a	n/a
3	d	5	2.55	3.19	3.19	n/a	2.60
4	e	Naphthalene	2.99	n/a	n/a	n/a	n/a
5	f	1	2.65	3.35	3.35	n/a	n/a
6	g	5	n/a	n/a	n/a	n/a	n/a
7	h	7	2.55	3.24	3.24	2.65	2.65
8	i	6	2.42	3.08	3.08	n/a	n/a
9	j	5	2.36	3.08	3.08	n/a	2.50
10	k	4	2.46	3.35	3.35	n/a	n/a
11	l	4	2.26	3.10	3.10	n/a	n/a
12	m	1	2.81	n/a	n/a	n/a	n/a
13	n	5	n/a	n/a	n/a	n/a	n/a
14	o	7	2.36	3.09	3.09	2.50	2.50
15	p	3	n/a	n/a	n/a	n/a	n/a
16	q	5	2.62	3.42	3.42	n/a	2.86
17	r	5	2.69	3.53	3.53	n/a	2.96
18	s	5	1.83	2.65	2.65	n/a	n/a
19	t	5	2.04	2.89	2.89	n/a	2.29
20	u	5	2.23	3.13	3.13	n/a	n/a
21	v	7	2.67	3.35	3.58	2.90	2.90
22	w	7	2.74	3.72	3.72	n/a	n/a
23	x	7	2.87	3.92	3.92	n/a	n/a
24	y	7	2.30	n/a	n/a	n/a	n/a
25	z	7	2.48	n/a	n/a	n/a	n/a

Table S3.3: Retention time of detected compounds (in minutes, continued).

		Numbering	1	4	6	7	8
Line Number	alphabet	position	tryptophan	CPA	Proviolacein	Deoxyviolacein	Violacein
0	a	0	203.0820	384.0984	326.0930	326.0930	342.0879
1	b	3	217.0977	n/a	n/a	n/a	n/a
2	c	3	217.0977	n/a	n/a	n/a	n/a
3	d	5	217.0977	412.1297	354.1242	354.1242	n/a
4	e	Naphthalene	214.0868	n/a	n/a	n/a	n/a
5	f	1	217.0977	412.1297	354.1242	354.1242	n/a
6	g	5	219.0770	n/a	n/a	n/a	n/a
7	h	7	217.0977	412.1297	354.1242	354.1242	n/a
8	i	6	221.0726	420.0796	n/a	362.0741	n/a
9	j	5	221.0726	420.0796	n/a	362.0741	n/a
10	k	4	217.0977	n/a	354.1242	354.1242	n/a
11	l	4	221.0726	n/a	n/a	n/a	n/a
12	m	1	220.0432	n/a	n/a	n/a	n/a
13	n	5	n/a	n/a	n/a	n/a	n/a
14	o	7	221.0726	420.0796	n/a	362.0741	n/a
15	p	3	n/a	n/a	n/a	n/a	n/a
16	q	5	237.0431	452.0205	n/a	394.0150	n/a
17	r	5	280.9926	539.9195	n/a	n/a	n/a
18	s	5	231.0770	n/a	n/a	n/a	n/a
19	t	5	228.0773	n/a	n/a	n/a	n/a
20	u	5	248.0671	474.0686	n/a	n/a	n/a
21	v	7	237.0431	452.0205	n/a	n/a	n/a
22	w	7	280.9926	539.9195	n/a	n/a	n/a
23	x	7	328.9787	635.8917	n/a	n/a	n/a
24	y	7	228.0773	n/a	n/a	n/a	n/a
25	z	7	248.0671	n/a	n/a	n/a	n/a

Table S3.4: Molecular ion mass calculated for detectable compounds.

		Numbering	9	10	11	12	13
Line Number	alphabet	position	CPA_hetero	Deoxyviolacein _hetero_1	Deoxyviolacein _hetero_2	Violacein _hetero_1	Violacein _hetero_2
0	a	0	n/a	n/a	n/a	n/a	n/a
1	b	3	n/a	n/a	n/a	n/a	n/a
2	c	3	n/a	n/a	n/a	n/a	n/a
3	d	5	398.1141	340.1086	340.1086	n/a	356.1035
4	e	Naphthalene	395.1032	n/a	n/a	n/a	n/a
5	f	1	398.1141	340.1086	340.1086	n/a	n/a
6	g	5	n/a	n/a	n/a	n/a	n/a
7	h	7	398.1141	340.1086	340.1086	356.1035	356.1035
8	i	6	402.0890	344.0835	344.0835	n/a	n/a
9	j	5	402.0890	344.0835	344.0835	n/a	360.0784
10	k	4	398.1141	340.1086	340.1086	n/a	n/a
11	l	4	402.0890	344.0835	344.0835	n/a	n/a
12	m	1	401.0596	n/a	n/a	n/a	n/a
13	n	5	n/a	n/a	n/a	n/a	n/a
14	o	7	402.0890	344.0835	344.0835	360.0784	360.0784
15	p	3	n/a	n/a	n/a	n/a	n/a
16	q	5	418.0595	360.0540	360.0540	n/a	376.0489
17	r	5	462.0089	404.0035	404.0035	n/a	419.9984
18	s	5	412.0934	354.0879	354.0879	n/a	n/a
19	t	5	409.0937	351.0882	351.0882	n/a	367.0831
20	u	5	429.0835	371.0780	371.0780	n/a	n/a
21	v	7	418.0595	360.0540	360.0540	376.0489	376.0489
22	w	7	462.0089	404.0035	404.0035	n/a	n/a
23	x	7	509.9951	451.9896	451.9896	n/a	n/a
24	y	7	409.0937	n/a	n/a	n/a	n/a
25	z	7	429.0835	n/a	n/a	n/a	n/a

Table S3.4: Molecular ion mass calculated for detectable compounds (continued).

		Numbering	1	4	6	7	8
Line Number	alphabet	position	tryptophan	CPA	Proviolacein	Deoxyviolacein	Violacein
0	a	0	203.0835	384.0977	326.0930	326.0936	342.0884
1	b	3	217.0963	n/a	n/a	n/a	n/a
2	c	3	217.0965	n/a	n/a	n/a	n/a
3	d	5	217.0971	412.1295	354.1237	354.1246	n/a
4	e	Naphthalene	214.0865	n/a	n/a	n/a	n/a
5	f	1	217.0977	412.1296	354.1239	354.1239	n/a
6	g	5	219.0768	n/a	n/a	n/a	n/a
7	h	7	217.0986	412.1298	354.1242	354.1237	n/a
8	i	6	221.0721	420.0792	n/a	362.0741	n/a
9	j	5	221.0721	420.0792	n/a	362.0739	n/a
10	k	4	217.0978	n/a	354.1242	354.1248	n/a
11	l	4	221.0722	n/a	n/a	n/a	n/a
12	m	1	220.0435	n/a	n/a	n/a	n/a
13	n	5	n/a	n/a	n/a	n/a	n/a
14	o	7	221.0726	420.0786	n/a	362.0739	n/a
15	p	3	n/a	n/a	n/a	n/a	n/a
16	q	5	237.0442	452.0204	n/a	394.0161	n/a
17	r	5	280.9925	539.9194	n/a	n/a	n/a
18	s	5	231.0774	n/a	n/a	n/a	n/a
19	t	5	228.0773	n/a	n/a	n/a	n/a
20	u	5	248.0666	474.0695	n/a	n/a	n/a
21	v	7	237.0438	452.0207	n/a	n/a	n/a
22	w	7	280.9933	539.9197	n/a	n/a	n/a
23	x	7	328.9799	635.8907	n/a	n/a	n/a
24	y	7	228.0777	n/a	n/a	n/a	n/a
25	z	7	248.0671	n/a	n/a	n/a	n/a

Table S3.5: Molecular ion mass observed for detectable compounds.

		Numbering	9	10	11	12	13
Line Number	alphabet	position	CPA_hetero	Deoxyviolacein _hetero_1	Deoxyviolacein _hetero_2	Violacein _hetero_1	Violacein _hetero_2
0	a	0	n/a	n/a	n/a	n/a	n/a
1	b	3	n/a	n/a	n/a	n/a	n/a
2	c	3	n/a	n/a	n/a	n/a	n/a
3	d	5	398.1141	340.1085	340.1085	n/a	356.1034
4	e	Naphthalene	395.1033	n/a	n/a	n/a	n/a
5	f	1	398.1140	340.1083	340.1083	n/a	n/a
6	g	5	n/a	n/a	n/a	n/a	n/a
7	h	**7**	398.1140	340.1081	340.1089	356.1035	356.1035
8	i	6	402.0891	344.0835	344.0835	n/a	n/a
9	j	5	402.0887	344.0835	344.0835	n/a	360.0784
10	k	4	398.1140	340.1089	340.1089	n/a	n/a
11	l	4	402.0887	344.0835	344.0835	n/a	n/a
12	m	1	401.0584	n/a	n/a	n/a	n/a
13	n	5	n/a	n/a	n/a	n/a	n/a
14	o	7	402.0890	344.0833	344.0833	360.0786	360.0786
15	p	3	n/a	n/a	n/a	n/a	n/a
16	q	5	418.0595	360.0542	360.0542	n/a	376.049
17	r	5	462.0092	404.0031	404.0031	n/a	419.0084
18	s	5	412.0930	354.0882	354.0882	n/a	n/a
19	t	5	409.0937	351.0888	351.0888	n/a	n/a
20	u	5	429.0833	371.0784	371.0784	n/a	n/a
21	v	7	418.0596	360.0542	360.0543	376.0490	376.0490
22	w	7	462.0088	404.0030	404.0030	n/a	n/a
23	x	7	509.9951	451.9893	451.9893	n/a	n/a
24	y	7	409.0938	n/a	n/a	n/a	n/a
25	z	7	429.0833	n/a	n/a	n/a	n/a

Table S3.5: Molecular ion mass observed for detectable compounds (continued).

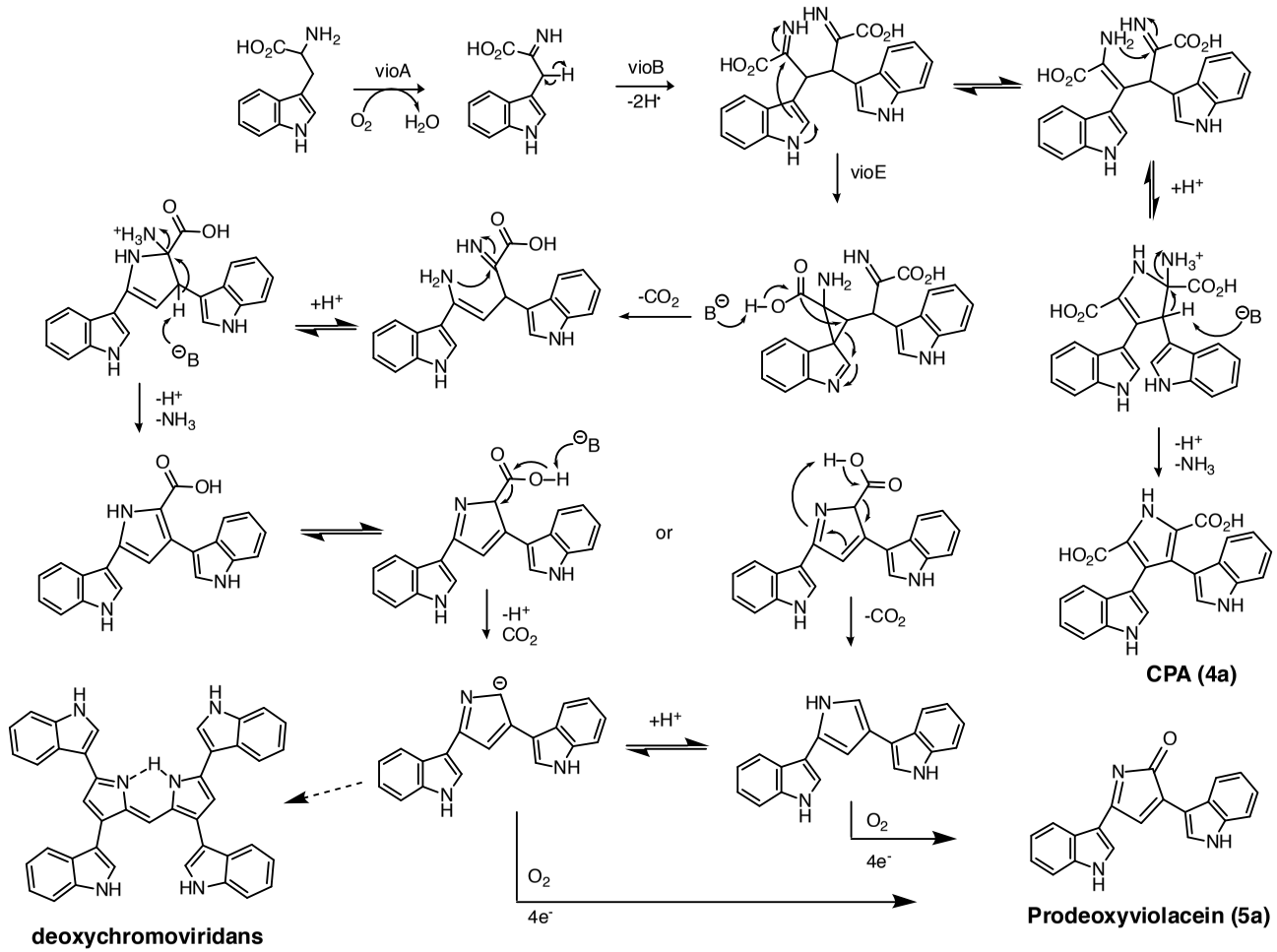


Figure S3.11: Mechanism for vioE-mediated divergence of prodeoxyviolacein (**5**) from Chromopyrrolic Acid (CPA, **4**), adapted from Scheme 2 of Balibar and Walsh., 2006.

Elemental Composition Report

Single Mass Analysis
Tolerance = 10.0 mDa / DBE: min = -1.5, max = 50.0
Element prediction: Off
Number of isotope peaks used for i-FIT = 2

Monoisotopic Mass, Even Electron Ions
1331 formula(e) evaluated with 37 results within limits (up to 50 closest results for each mass)
Elements Used:
C: 0-40 H: 0-80 N: 0-5 O: 0-20 F: 0-3

Minimum:				-1.5				
Maximum:		10	10	50				
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Norm	Conf(%)	Formula
362.0739	362.0737	0.2	0.6	11.5	250.4	5.944	0.26	C14 H12 N5 O7
	362.0741	-0.2	-0.6	16.5	252.5	7.979	0.03	C20 H10 N3 O2 F2
	362.0735	0.4	1.1	2.5	255.6	11.078	0	C10 H17 N O12 F
	362.0746	-0.7	-1.9	-1.5	256.6	12.127	0	C7 H18 N O13 F2
	362.0748	-0.9	-2.5	7.5	253.2	8.677	0.02	C11 H13 N5 O8 F
	362.073	0.9	2.5	20.5	254.5	9.966	0	C23 H9 N3 O F
	362.0753	-1.4	-3.9	12.5	244.6	0.123	88.41	C17 H11 N3 O3 F3
	362.0723	1.6	4.4	6.5	254	9.532	0.01	C13 H16 N O11
	362.0718	2.1	5.8	24.5	255.7	11.245	0	C26 H8 N3
	362.076	-2.1	-5.8	3.5	254.8	10.301	0	C8 H14 N5 O9 F2
	362.0712	2.7	7.5	8.5	252.9	8.383	0.02	C12 H11 N5 O5 F3
	362.0771	-3.2	-8.8	-0.5	256.6	12.061	0	C5 H15 N5 O10 F3
	362.0777	-3.8	-10.5	15.5	252	7.515	0.05	C19 H12 N3 O5
	362.0701	3.8	10.5	12.5	249.3	4.788	0.83	C15 H10 N5 O4 F2
	362.0699	4	11	3.5	255.6	11.092	0	C11 H15 N O9 F3
	362.0781	-4.2	-11.6	20.5	256	11.546	0	C25 H10 N F2
	362.0695	4.4	12.2	-1.5	256.6	12.112	0	C5 H17 N3 O14 F
	362.0788	-4.9	-13.5	11.5	248.6	4.137	1.6	C16 H13 N3 O6 F
	362.0689	5	13.8	16.5	251.3	6.786	0.11	C18 H9 N5 O3 F
	362.0687	5.2	14.4	7.5	253.8	9.339	0.01	C14 H14 N O8 F2
	362.0793	-5.4	-14.9	16.5	254.8	10.31	0	C22 H11 N O F3
	362.0795	-5.6	-15.5	2.5	256.3	11.755	0	C7 H16 N5 O12
	362.0683	5.6	15.5	2.5	256.2	11.735	0	C8 H16 N3 O13
	362.0678	6.1	16.8	20.5	254.5	10.004	0	C21 H8 N5 O2
	362.08	-6.1	-16.8	7.5	254.1	9.554	0.01	C13 H14 N3 O7 F2
	362.0676	6.3	17.4	11.5	247	2.523	8.02	C17 H13 N O7 F
	362.0807	-6.8	-18.8	-1.5	258.1	13.615	0	C4 H17 N5 O13 F
	362.0811	-7.2	-19.9	3.5	256.2	11.663	0	C10 H15 N3 O8 F3
	362.0665	7.4	20.4	15.5	253.7	9.173	0.01	C20 H12 N O6
	362.0817	-7.8	-21.5	19.5	256.5	11.99	0	C24 H12 N O3
	362.0659	8	22.1	-0.5	257.7	13.199	0	C6 H15 N3 O11 F3
	362.0654	8.5	23.5	17.5	253.6	9.116	0.01	C19 H7 N5 F3
	362.0829	-9	-24.9	15.5	255.1	10.578	0	C21 H13 N O4 F
	362.0647	9.2	25.4	3.5	257.1	12.587	0	C9 H14 N3 O10 F2
	362.0643	9.6	26.5	-1.5	259.5	15.023	0	C3 H16 N5 O15
	362.0836	-9.7	-26.8	6.5	255.8	11.292	0	C12 H16 N3 O10
	362.064	9.9	27.3	12.5	249.7	5.19	0.56	C18 H11 N O4 F3

Figure S3.12: Element composition analysis of compound **7o**. Highlighted formula matches with required element composition.

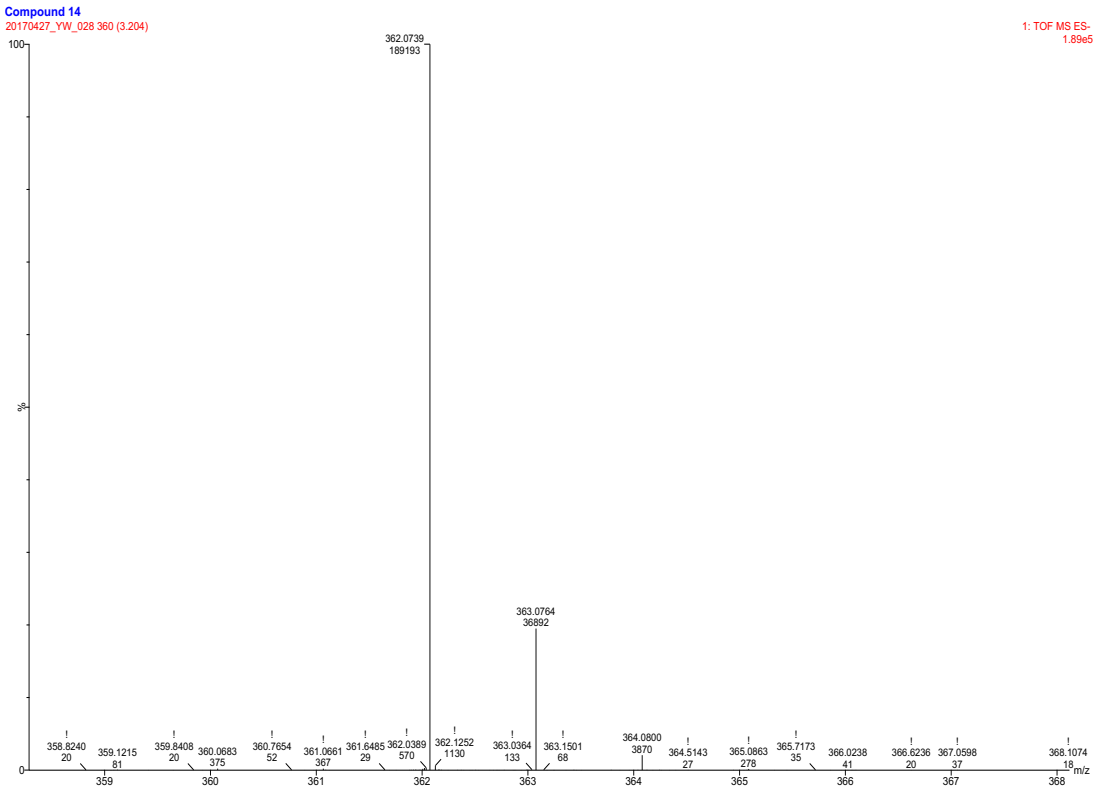


Figure S3.12: Element composition analysis of compound **7o** (continued).

Elemental Composition Report

Single Mass Analysis
Tolerance = 10.0 mDa / DBE: min = -1.5, max = 50.0
Element prediction: Off
Number of isotope peaks used for i-FIT = 2

Monoisotopic Mass, Even Electron Ions
375 formula(e) evaluated with 9 results within limits (up to 50 closest results for each mass)
Elements Used:
C: 0-40 H: 0-80 N: 0-5 O: 0-20

Minimum:				-1.5				
Maximum:		10	10	50				
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Norm	Conf(%)	Formula
354.0882	354.0884	-0.2	-0.6	-0.5	69.4	1.207	29.92	C8 H20 N O14
	354.0879	0.3	0.8	17.5	70.2	2.003	13.49	C21 H12 N3 O3
	354.0897	-1.5	-4.2	4.5	69.7	1.496	22.4	C9 H16 N5 O10
	354.0919	-3.7	-10.4	21.5	70.8	2.63	7.2	C26 H12 N O
	354.0838	4.4	12.4	13.5	70.7	2.484	8.34	C16 H12 N5 O5
	354.0937	-5.5	-15.5	8.5	70.9	2.678	6.87	C14 H16 N3 O8
	354.0825	5.7	16.1	8.5	70.9	2.731	6.51	C15 H16 N O9
	354.0978	-9.6	-27.1	12.5	72.1	3.879	2.07	C19 H16 N O6
	354.0785	9.7	27.4	4.5	71.7	3.445	3.19	C10 H16 N3 O11

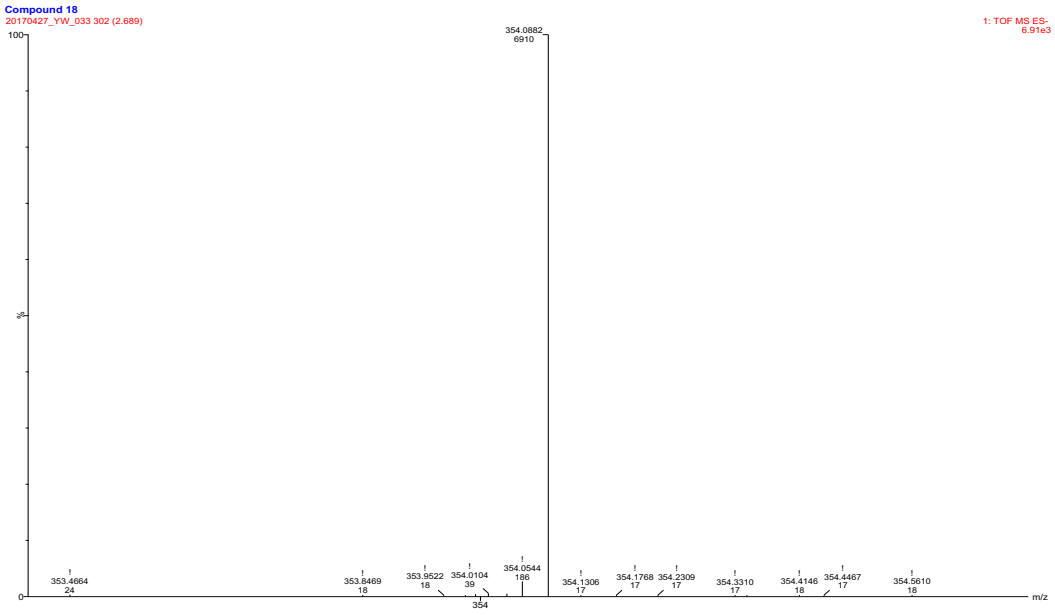


Figure S3.13: Element composition analysis of compound **10s** or **11s**. Highlighted formula matches with required element composition.

Chapter 4

CONCLUSION AND FUTURE DIRECTIONS

This work has demonstrated the cell-free TX-TL system as a reliable and convenient platform for the engineering of biosynthetic pathways. The TX-TL system is capable of expressing multiple pathway enzymes and simultaneously maintaining an active metabolism. The lack of cell membrane enables high-throughput screening of multiple pathway combinations, rapid tuning of protein expression levels, and efficient substrate scope studies. This work presents three different biosynthetic pathways for exploring various parts of the metabolic network in the TX-TL system. The production of 2,3-butanediol, 1,4-butanediol, and violacein showed that the TX-TL system is capable of producing compounds of different chemical spaces. These three pathways are cofactor dependent, and the production of final pathway metabolites suggests active cofactor metabolism. Chapter 2 and Chapter 3 have both demonstrated the practicality of TX-TL. Chapter 2 has shown the mapping between the cell-free TX-TL system and the small-scale *in vivo* system, in which case, similar findings are drawn from the design space exploration of a metabolic pathway in both systems. Chapter 3 has indirectly shown the comparison between the cell-free TX-TL system and the *in vitro* system made with purified proteins, in which case, findings drawn from the TX-TL system agrees with previous findings drawn from the *in vitro* system. The cell-free TX-TL system can become a new tool that bridges the whole cell system and the purified protein assay, which offers an alternative platform for the engineering of biological systems.

The cell-free TX-TL system allows for analyzing biosynthetic pathways in modules, which enables engineering biosynthetic pathway with a real systems approach, rather than trial and error. The system provides an alternative means to identify metabolic bottlenecks, study reaction mechanism, and characterize pathway dynamics. A biosynthetic pathway can be easily de-compartmentalized and reassembled by manipulating the DNA materials added to the system. Substrates or intermediates can be added to test for enzyme expression and enzyme activity, and individual components of natural biological systems can also be removed for investigations. The direct access to cellular components also reduces the timeline for design-build-test. By adding linear DNA in the cell-free TX-TL system for prototyping pathways, results of enzyme expression levels and metabolic outputs can be obtained in one day.

We envision that the cell-free TX-TL system will become the standard research and development platform for engineering biological systems in the future. Similar to the wind tunnels used for aeronautic research, the cell-free TX-TL system can be reconfigured for different prototyping applications for engineering biosynthetic pathways. This work has mainly focused on the industrial relevance of TX-TL, and the capabilities of TX-TL have not been fully investigated. However, the cell-free TX-TL system can be further explored by making extracts from different organisms,[111-112] reformulating buffer to reroute metabolic network,[66, 113] and identifying dimensionless parameters that capture biochemical dynamics.[114-116] The gap between TX-TL and the whole cell *in vivo* system or the gap between TX-TL and the purified protein *in vitro* system can be filled in with mathematical

models and computational simulations. The biomolecular breadboard TX-TL can also help identify dynamics of biological systems by providing access for measuring key parameters.

Engineering the cell-free TX-TL system for different applications can help TX-TL better complement bioengineering research. The ingredients that are added into the currently TX-TL system can be expensive and potentially non-essential. Reformulating buffer can develop a metabolic network similar to the one *in vivo*, prolong system lifetime, and most likely reduce costs. All of which can further enhance the industrial relevance of the cell-free TX-TL system. Further engineering and development efforts towards the investigation of the viscoelastic properties of TX-TL can also enable the commercialization. Research and development efforts for optimizing mixing and will guide the efforts for large-scale synthesis, automated droplet formation, and screen assay with microfluidics.

Challenges of the cell-free TX-TL system include quality control on a few levels. At the dawn of cell-free extract production in the Murray Lab, batch to batch variability is very prominent. However, the variability can be reduced by implementing effective cell-growth control, using the same protocol, and maintain procedure consistency. Variability on the buffer used can be problematic. Current protocol calls for small amounts of unstable cofactors, which can be improved through reformulation and reliable measurement methods. Quality control of ingredients added to the cell-free TX-TL system will also improve reproducibility of experimental results. For example, the purity of DNA added into the system can impact protein synthesis, and quality control of the DNA added can ensure consistent results.

The opportunities for the cell-free TX-TL system include exploring unfamiliar biochemical environments and expanding the biochemical space for biosynthesis. Chapter 3 demonstrates the production of natural products that originate from an organism that can cause human infection.[117-118] With the help of transcriptomics and proteomics, the investigation of heterologous biosynthetic pathways in *E.coli* based TX-TL can be a much safer alternative for natural product discovery. Similarly, extracts of invasive organisms can be made to investigate its metabolic network in a safer environment. Cell growth can be eliminated in the TX-TL platform, where an active biochemical network is still maintained. The potential toxicity effect of the produced metabolites may be reduced since cell growth has been eliminated. For product characterization, the production of target metabolites can be scaled up using a continuous exchange system, which also extends the lifetime of TX-TL.

Appendix A

DESIGN SPACE EXPLORATION OF THE 2,3-BUTANEDIOL PATHWAY

ABSTRACT

This work reports the production of meso-2,3-butanediol (2,3-BDO) in the cell-free transcription-translation (TX-TL) system. This work showcases the first example of simultaneous enzyme expression and chemical synthesis from non-native pathways in the cell-free TX-TL systems in the Murray Lab. 1g/L of 2,3-BDO is observed after overnight TX-TL reactions. The production of 2,3-BDO validates the active reduction of NAD^+ in the TX-TL system. Combinations of various concentrations of plasmids encoding individual pathway enzymes are tested for improved metabolite production rate. We find that increasing the concentration of plasmid encoding the first enzyme of the pathway increases the target production rate by two-fold.

INTRODUCTION

The cell-free transcription-translation (TX-TL) system is becoming a platform for the prototyping of biological systems. Specifically, the cell-free TX-TL system has been used to prototype biological circuits. Although other cell-free systems have been applied for protein synthesis or metabolite production, simultaneous protein synthesis and metabolite production from a non-native pathway have not been reported before this work. The meso-2,3-butanediol (2,3-BDO) is used to test the capabilities of the cell-free TX-TL system. 2,3-

BDO has applications in pharmaceutical, agrochemical, and fine chemical industries.[119-120] Studies have shown significant promise for the commercial production of bio-based 2,3-BDO.[121-122] Enzymes of the pathway have been previously characterized *in vitro*.[123-125] The 2,3-BDO pathway consists of three enzymes (shown in Figure A.1): acetolactate synthase (alsS) converts two pyruvate molecules into (S)-2-acetolactate and carbon dioxide, acetolactate decarboxylase (alsD) converts (S)-2-acetolactate into (R)-3-acetoin (ACT), and 2,3-BDO dehydrogenase (budC) converts ACT to 2,3-BDO with the help of NADH. In this work, we explore the design space of the 2,3-BDO pathway by measuring metabolite production rates of TX-TL reactions with varying enzyme expression levels, which is achieved by varying the concentration of plasmids encoding individual pathway enzymes. We also change the concentrations of cofactor and substrates to understand dynamics of the cell-free TX-TL system.

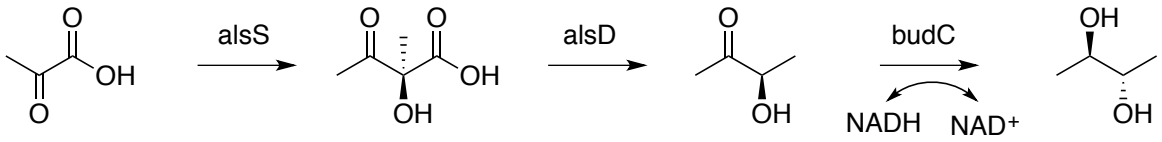


Figure A.1: The meso-2,3-butanediol biosynthetic pathway.

RESULTS AND DISCUSSION

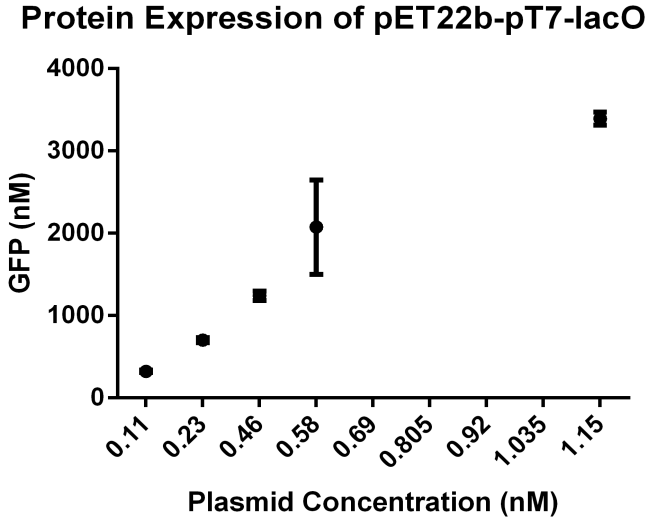
We first test different reaction conditions for improved production of 2,3-BDO in TX-TL using plasmids from previous studies.[124] Plasmid encoding T7 polymerase (plT7) and IPTG are added to TX-TL reactions to induce protein synthesis. The effect of plT7 on protein

expression in cell-free TX-TL is characterized using a plasmid encoded with a T7 promoter followed by GFP. The final concentration of GFP produced in TX-TL is linearly dependent on the plasmid concentration added to reactions in the range between 0.1 to 1 nM (shown in Figure A.2a). The effect of plT7 and IPTG on metabolite production is shown in Figure A.2b. We specifically test 0.5 mM, 2.5 mM, and 5 mM of IPTG and 0.2 nM, 0.4 nM, 0.6 nM, and 0.7 nM of plT7. We find 0.6 nM of plT7 and 0.5 mM IPTG is enough for the required enzyme expression that leads to metabolite production, which matches with the value shown in previous studies.[65] The production level of 2,3-BDO stays around 1 g/L or 11 mM. Such translates to at least 11 mM of NADH is reduced in TX-TL. We test the effect of adding extra NAD^+ to the cell-free TX-TL system and find that extra NAD^+ does not impact BDO production positively or negatively (data not shown). We keep the concentration of NAD^+ at 0.33 mM for the rest of the reactions.

TX-TL reactions with varying enzyme expression levels are achieved by modulating plasmid DNA concentrations. Segments of genes encoding individual enzymes are cloned under the same promoter (T7-lacO) and RBS from the original vector. Plasmid encoding individual enzymes are added into TX-TL reactions to test for metabolite production. TX-TL reactions added with multiple plasmids are compared to TX-TL reactions added with the original plasmid encoding all pathway enzymes (shown in Figure A.3). The production level of 2,3-BDO is very similar in both reactions. The rate of metabolite production in TX-TL is examined. Time course measurements of TX-TL reactions added with different plasmid ratios are conducted for up to 4 hours. The TX-TL system can be resource limited. The rate of protein production is approximately constant within the first 4 hours of TX-TL reactions

for plasmids with strong promoters and of concentrations lower than 1 nM.[126] The production rate of metabolite within the first 4 hours can reveal the effect of plasmid concentration on protein production and possibly metabolite production.

a.



b.

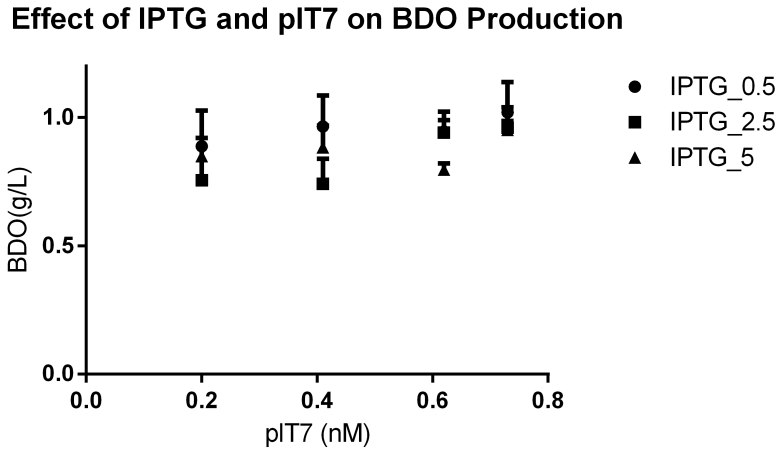


Figure A.2: The effect of T7 polymerase and IPTG on 2,3-BDO production. a. GFP Production under T7-lacO promoter. b. The effect of IPTG and plasmid encoding T7 polymerase on the BDO production. BDO concentration above is measured at the end of 12-hour TX-TL reactions. Series with circle represents result from reactions with 0.5 mM of IPTG, series with square represents result from reactions with 2.5 mM of IPTG, and series with triangle represents result from reactions with 5 mM of IPTG.

The original plasmid design may not have balanced enzyme expression for metabolite production. Initial testing of the original plasmid indicates significant accumulation of the intermediate ACT **(**shown in Figure A.3**)**, which suggests that the expression of enzyme budC is much lower than enzyme alsS or enzyme alsD. There is also no measurement of 2,3-BDO until two hours into the TX-TL reactions, which suggests the original plasmid construct may limit budC production. The original plasmid contains genes encoding pathway enzymes in a polycistronic manner. The low budC expression might be due to the mRNA secondary structure or the lack of internal RBS optimization of the original plasmid design.[127] Monocistronic design or adjusting RBS strength can help balance enzyme expression for improved target metabolite production.

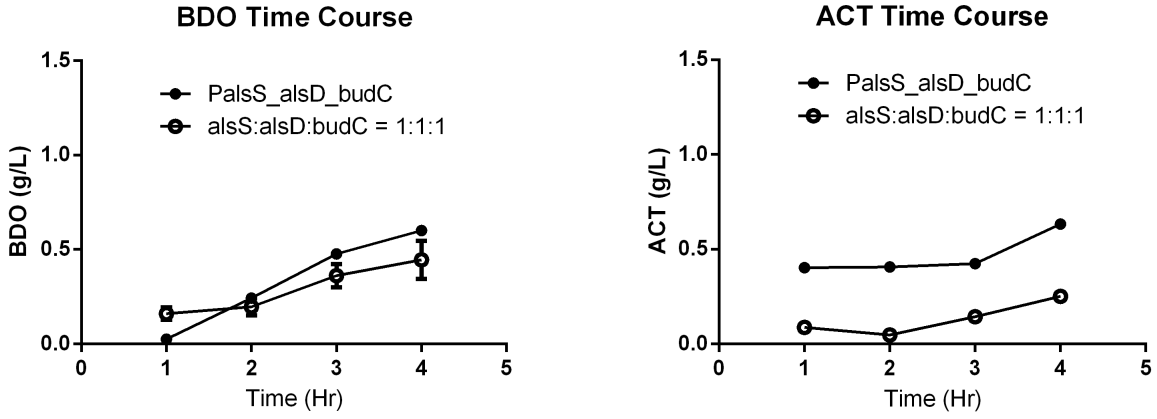


Figure A.3: Metabolite production in TX-TL reactions. Above shows the first four hours of time course measurement of BDO (left) and ACT (right) from TX-TL reaction with 1 nM of original plasmid with alsS, alsD, and budC arranged in polycistronic fashion (closed circle) and TX-TL reaction with 1 nM of each plasmid encoding a pathway enzyme (open circle).

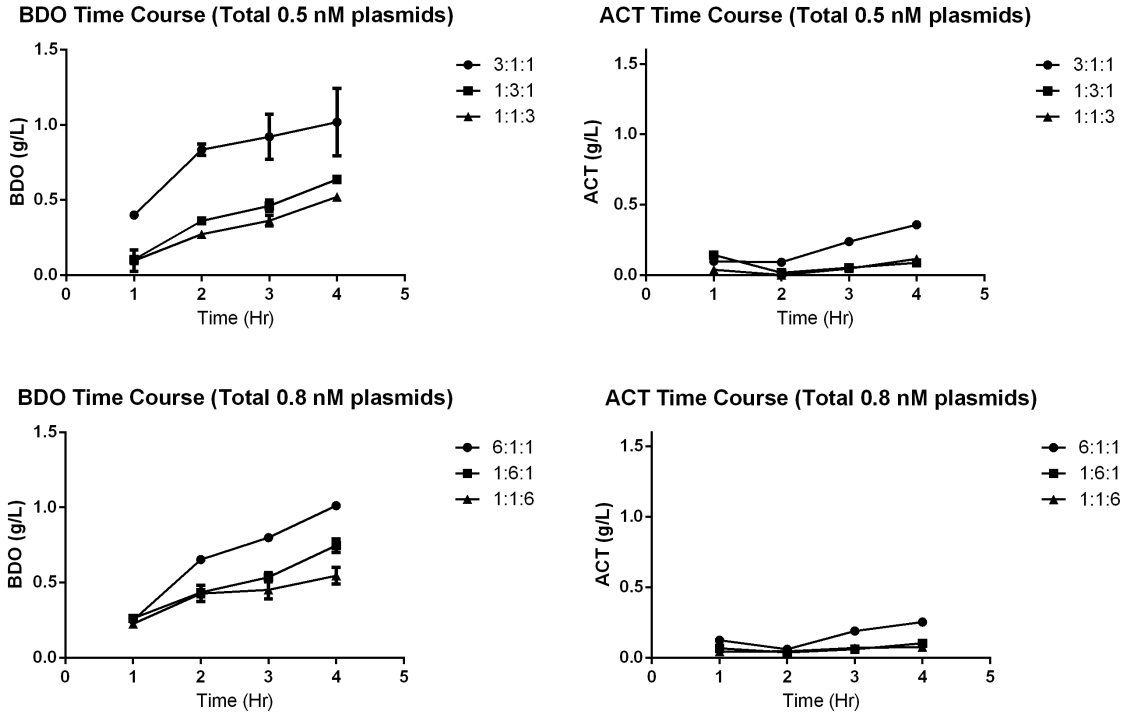


Figure A.4: Modulating enzyme expression levels for improved 2,3-BDO production in TX-TL. Top two graphs show BDO (left) and ACT (right) time course measurement for plasmid ratio of alsS:alsD:budC = 3:1:1 (circle), 1:3:1(square), and 1:1:3 (triangle), and total DNA concentration of 0.5 nM. Bottom two graphs show BDO (left) and ACT (right) time course measurement for plasmid ratio of alsS:alsD:budC = 6:1:1 (circle), 1:6:1 (square), 1:1:6 (triangle), and total DNA concentration of 0.8 nM.

We tested the effect of each enzyme on the production of 2,3-BDO by varying concentration ratio of plasmid encoding the particular enzyme. Experimental data indicates that the addition of plasmid encoding enzyme alsS significantly increases the production of 2,3-BDO. Increasing the concentration of plasmid encoding enzyme alsD or budC does not substantially increase the production rate of 2,3-BDO or ACT. Figure A.4 shows that increasing the ratio of plasmid encoding alsS increases the production rate of 2,3-BDO. We hypothesize that the metabolite production is limited by the expression of alsS. The

nucleotide sequence length for alsS is twice as long as the ones for alsD and budC. The transcription and the translation rate for alsS are relatively lower as a result of its sequence length (A.E.1). Furthermore, the kinetic parameters from previous studies also indicate that the catalytic rate of alsS is lower than alsD or budC (data shown below),[125] which may hint that alsS is the rate-limiting enzyme. The highest production rate observed was 0.4gBDO/L/hr with plasmid ratio of alsS:alsD:budC = 3:1:1.

A mathematical model that simulates the 2,3-BDO pathway dynamics in the cell-free TX-TL system is presented. Equation A.E.1 describes the dynamic of mRNAs (m_i) in the system, and i corresponds to the pathway enzymes alsS, alsD, and budC:

$$\frac{dm_i}{dt} = \frac{\frac{k_{TX}}{N_{m,i}} k_{m,f} P s_i R}{\left(\frac{k_{TX}}{N_{m,i}} + k_{m,r}\right) + R} - \frac{m_i}{\tau_m}, \quad \forall\ t > \tau_0 \tag{A.E.1}$$

where k_{TX} represents the transcription rate (previously measured to be 1 ± 0.5 mRNA nucleotide per second[128]), which is at least one order magnitude smaller than *in vivo*. $N_{m,i}$ represents the length of a given mRNA, 1686 for alsS, 768 for alsD, and 771 for budC. Ps_i represents the concentration of plasmid encoding pathway enzyme i. R represents the concentration of mRNA polymerase (which is approximately 30nM[128]). τ_m accounts for the exponential decay lifetime (which is about 12 minutes). $k_{m,f}$ and $k_{m,r}$ are unknown, but $K_{TX} = k_{m,f}/(k_{TX}/N_{m,i}+k_{m,r})$ was empirically measured to be approximately 2 nM.[128] τ_0 represents a time delay for mRNA production, which is about 15 min.[128] Equation A.E.2 describes the dynamic of enzymes (E_i) in the system, and i corresponds to the pathway enzymes alsS, alsD, and budC:

$$\frac{dE_i}{dt} = \frac{\frac{k_{TL}}{N_{p,i}} k_{E,f} m_i b}{\left(\frac{k_{TL}}{N_{p,i}} + k_{E,r}\right) + b} - kd \cdot Ei \quad \text{(A.E.2)}$$

where k_{TL} represents the translation rate, which has a lower bound of 4 amino acids per second[128] and is also less than the value *in vivo*. $N_{p,i}$ represents the length of a given enzyme, 562 for alsS, 256 for alsD, and 257 for budC. b represents the concentration of ribosomes (> 30nM[128]). $kd \cdot E_i$ represents the enzyme degradation rate (approximately 15 nM/min). $k_{E,f}$ and $k_{E,r}$ are unknown, but $K_{TL} = k_{E,f}/(k_{TL}/N_{E,i}+k_{E,r})$ can be approximated to be less than 0.5 nM based on previous empirical data.[128] Equation A.E.3 describes the pyruvate (*Pyr*) consumption in the system:

$$\frac{dPyr}{dt} = -\frac{kc,_P E_{alsS} Pyr}{K_{m,p}+Pyr} - kPyr \quad \text{(A.E.3)}$$

where $k_{c,p}$ and $K_{m,p}$ represents the Michaelis-Menten constants obtained (121/s and 13.6 mM respectively[125]), k is a consumption rate added to simulate the consumption of pyruvate for TCA cycle. Since the concentration of enzyme *alsS* would remain zero for about 20 min, the excess consumption accounts for the decrease of pyruvate concentration during time delay for mRNA production. Equation A.E.4 describes the acetolactate (*Actl*) dynamics in the system:

$$\frac{dActl}{dt} = -\frac{kc,_a E_{alsD} Actl}{K_{m,a}+Actl} + \frac{kc,_p E_{alsS} Pyr}{K_{m,p}+Pyr} \quad \text{(A.E.4)}$$

where $k_{c,a}$ and $K_{m,a}$ represents the Michaelis-Menten constants obtained (4000/min and 1.3 mM respectively[129]). Equation A.E.5 and A.E.6 describes the 3-acetoin (*Act*) and meso-2,3-butanediol (*BDO*) dynamics in the TX-TL system:

$$\frac{dAct}{dt} = -\frac{kc_{,b}E_{budC}Act}{K_{m,b}+Act} + \frac{kc_{,a}E_{alsD}Actl}{K_{m,a}+Actl} \quad \text{(A.E.5)}$$

$$\frac{dBDO}{dt} = \frac{kc_{,b}E_{budC}Act}{K_{m,b}+Act} \quad \text{(A.E.6)}$$

where $k_{c,b}$ and $K_{m,b}$ represents the Michaelis-Menten constants obtained (58/s and 0.85 mM[124]). Note that $k_{c,b}$ and $K_{m,b}$ were measured with the presence of significant amount of NADH, but NADH is a limiting factor for *in vivo* systems.[130] $k_{c,b}$ and $K_{m,b}$ were also measured with the assumption that the forward rate and reverse rate for the production of 2,3-BDO is in equilibrium. The result of the mathematical model is very different from data collected experimentally (shown in Figure A.5). The model predicts the production of 2,3-BDO starts to plateau in less than two hours, and the consumption of the acetoin completes by the end of two hours. The difference between the simulation and experimental data might be because the model has not taken into account resource limitations of the TX-TL system. We hypothesize that the cofactor NADH must have been slowly released over time in the TX-TL, and there can be other competing enzymes in the TX-TL that also utilizes NADH for other background reactions.

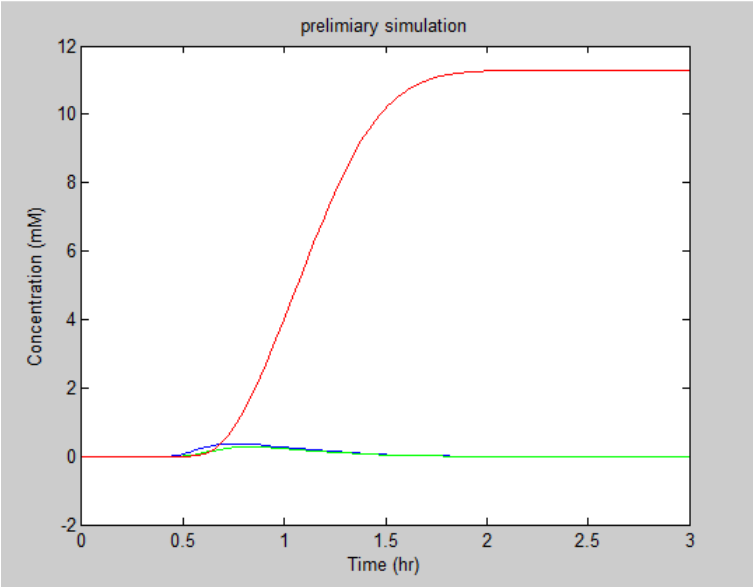


Figure A.5: Simulation of the 2,3-BDO pathway in TX-TL. Red represents the concentration of 2,3-BDO, blue represents the concentration of acetolactate, and green represents the concentration of acetoin.

MATERIALS AND METHODS

Cell-Free Expression Preparation and Execution

Preparation of the cell-free TX-TL expression system was done according to protocols previously described by colleagues,[33] resulting in extract with conditions: 8.9−9.9 mg/mL protein, 4.5−10.5 mM Mg-glutamate, 40−160 mM K-glutamate, 0.33−3.33 mM DTT, 1.5 mM each amino acid except leucine, 1.25 mM leucine, 50 mM HEPES, 1.5 mM ATP and GTP, 0.9 mM CTP and UTP, 0.2 mg/mL tRNA, 0.26 mM CoA, 0.33 mM NAD+, 0.75 mM cAMP, 0.068 mM folinic acid, 1 mM spermidine, 30 mM 3-PGA, 2% PEG-8000. When possible, inducers such as IPTG, pyruvate (60 mM), and plasmids were added to a mix of extract and buffer to ensure uniform distribution. TX-TL reactions were conducted in PCR tubes and kept at 29°C with incubation in PCR machine. The temperature was chosen for

maximum protein production in the cell-free system. Time course measurements were drawn from the same tube every hour, and each tube contains at least 20 ul reaction volumes. At the end of each reaction, 9 volume of methanol was added to precipitate proteins and terminate reactions. The reaction mixture was centrifuged for 15 min at 14000 rpm. The supernatant is taken for compound detection using GC-MS, and the protein precipitate is analyzed with SDS-PAGE gel.

Gas Chromatography Mass Spectrometry (GC-MS) Method

Supernatant from TXTL reaction mixture was analyzed with GC-MS (Hewlett Packard 6890 GC System and 5973 Mass Selective Detector) equipped with an Innowax column (Agilent, 30 m x 0.25mm ID, 0.25 um film thickness). A double gooseneck liner (Restek SKY Liner, splitless double taper gooseneck, 4 mm x 6.5 x 78.5) was used to present a minimally reactive surface to improve 2,3-BDO response. The GC oven program is 80°C for 1 min, 15 °C/min to 150 °C, then 60 °C/min to 230 °C, and held for 1 minute. The inlet temperature is 210 °C. Helium was used as carrier gas in constant flow mode at 0.9 ml/min. Injection volume was 1 ul, interface temperature was 250 °C, Quadrupole temperature was 126 °C, Acetoin and meso-2,3-butanediol were quantified with ion extraction at 45 m/z. The retention times are 3.7 min and 5.94 min, respectively.

Plasmid DNA and PCR Product Preparation

Plasmids used in this study were constructed using standard cloning procedures and maintained in a JM109 strain. PCR products were amplified using Pfu Phusion Polymerase

(New England Biolabs). Plasmids were miniprepped using QIAprep spin columns (Qiagen). All plasmids were processed at stationary phase. Before use in the cell-free reaction, both plasmids and PCR products underwent an additional PCR purification step using a QiaQuick column (Qiagen), which removed excess salt detrimental to TX-TL and were eluted and stored in 10 mM Tris-Cl solution, pH 8.5 at 4°C for short-term storage and −20°C for long-term storage.

ACKNOWLEDGEMENTS

We thank Jennifer Kay and Professor Michael Jewett from the Northwestern University for providing us plasmid for the 2,3-BDO pathway. We also thank them for verifying the production of 2,3-BDO using their liquid chromatography. We thank Zach Wickens and Professor Robert H. Grubbs for having us use their GC-FID to collect initial data for the project. We thank Dr. Nathan Dalleska and the Environmental Analysis Center for the support and assistance using GC-FID and GC-MS for data collection. Y.Y.W was supported by NIH/NRSA Training Grant 5 T32 GM07616 and DAPRA HR0011-12-C-0065 (DARPA/CMO).

Appendix B

HIGH-THROUGHPUT DESIGN SPACE EXPLORATION OF THE 1,4-BUTANEDIOL PATHWAY IN TX-TL

The goal of this work is to develop a high-throughput screening platform that utilizes the cell-free TX-TL system to explore the design space of a metabolic pathway. Exploring the optimal ratio of enzyme expressions in the cell-free TX-TL system may serve as an alternative to the *in vivo* system for improving metabolite production. Appendix A shows that the ratio of enzyme expressions can be modulated by varying concentration of plasmid encoding individual pathway enzymes. Here we explore the possibility of characterizing the 1,4-BDO pathway in TX-TL using a screening assay. It may be easy to predict the optimal ratio of pathway enzyme expression levels by analyzing the production rate of mRNA and protein mathematically. However, balancing enzyme expression levels of a complex metabolic pathway in a biological system with finite resources remains a challenge.[34]

We plan to work with a high-throughput platform to quickly scan many combinations of enzyme expression levels and cofactor concentrations for desired pathway functions. This work is in collaboration with the Abate Lab at UCSF to demonstrate design space exploration of a metabolic pathway on a high-throughput droplet-based microfluidic system. The system can generate droplets of aqueous liquid dispersed in an inert carrier oil, inject materials into droplet after droplet formation, and sort fluorescence-based molecules.[131] The system can create and sort 10^6 droplets in one day.

Genomatica has previously developed an assay enzyme that detects the production of 1,4-BDO by converting 1,4-BDO back to 4HB-aldehyde (Figure B.2). The conversion step also converts cofactor NADP to NADPH. The application of the assay enzyme to the TX-TL system will allow for high-throughput detection of 1,4-BDO. An NADPH assay kit can be applied to obtain fluorescent signals from each TX-TL reaction. The fluorescence signal is very critical for the system to function because of the droplet size. Figure B.1 shows the overall workflow of characterizing the 1,4-BDO pathway using the microfluidics system coupled with the cell-free TX-TL system. The left side of Figure B.1 represents different tubes containing DNA that encodes individual pathway enzymes, substrate, and TX-TL mixture. After TX-TL reactions with various conditions undergo incubation, an assay enzyme is added to consume 1,4-BDO and produce NADPH. The NADPH kit is applied to generate a corresponding fluorescent signal for measurement. The ability to collect a large amount of time-point data using a high-throughput system will enable system identification and therefore allow us to identify limiting factors for the production of 1,4-BDO.

From Figure B.3 to Figure B.4 shows the progression of understanding the logistics of applying an NADPH kit and the potential plan of executing a high-throughput screening platform. In Figure B.3, a process of heating and cooling is added to deactivate enzymes present in the TX-TL system, which can eliminate false positives and substrate competition between the added ADHX and ADH produced from TX-TL. It turns out that there are quite a few steps required for an NADPH kit to provide reliable fluorescent signals, which could be detrimental to droplet formation. Developing a high-throughput system with fluorescent proteins, click chemistry, or other simple fluorescent assays might be better alternatives.

This work also explores the use of assay enzyme ADHX for detecting 1,4-BDO in TX-TL. Two systems were developed: (1) fluorescence measurement using assay enzyme coupled with the NADPH kit and (2) absorbance measurement with assay enzyme. Figure B.5 shows the calibration curve for NADP and NADPH between 0 and 3 uM. Although the assay kit is specific for NADPH and not NADH, it is hard to differentiate NADPH or NADP (Figure B.5b). Figure B.6 shows that the signal of absorbance (a) and the signal of fluorescence against the concentration of 1,4-BDO.

Other venues of high-throughput screening include MALDI. Significant efforts will be needed to ensure the reliability of the MALDI system. Most importantly, the omission of PEG will be necessary for TX-TL to be a competitive platform against other cell-free options.

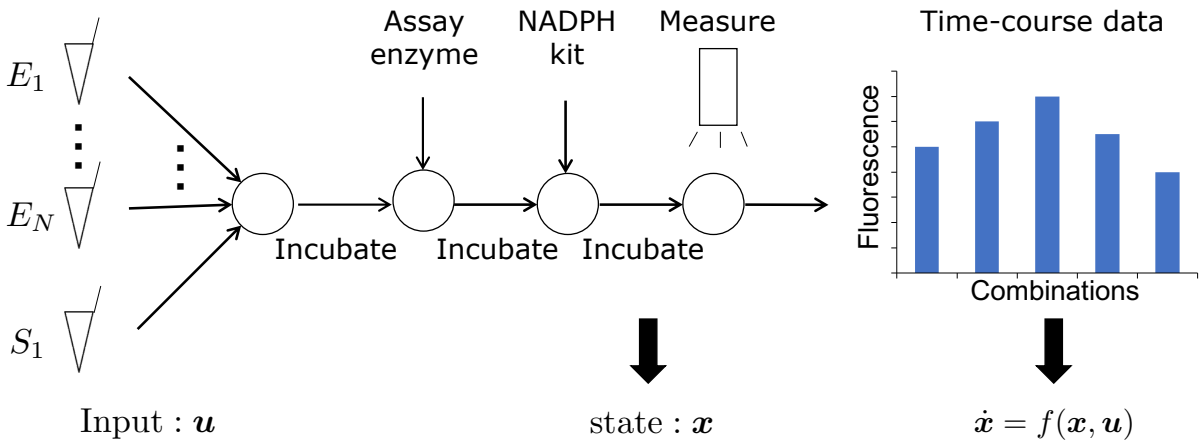


Figure B.1: High throughput design space exploration for system identification in TX-TL.

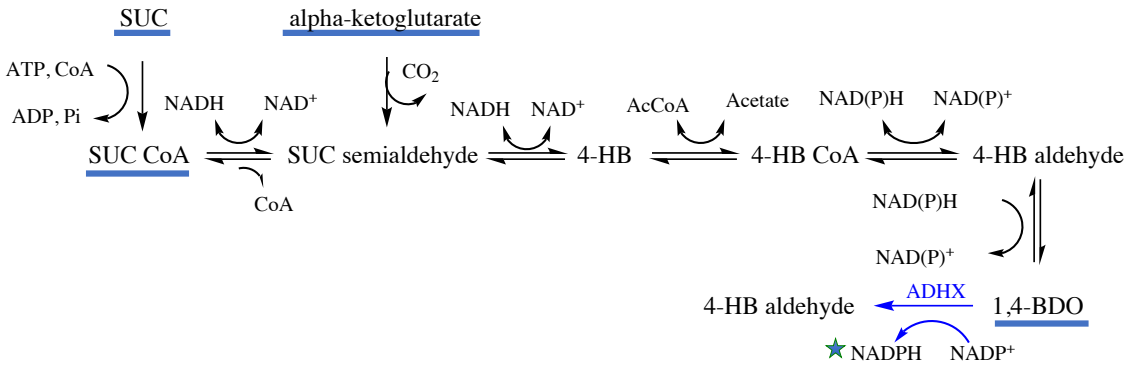


Figure B.2: 1,4-BDO pathway with a detection assay.

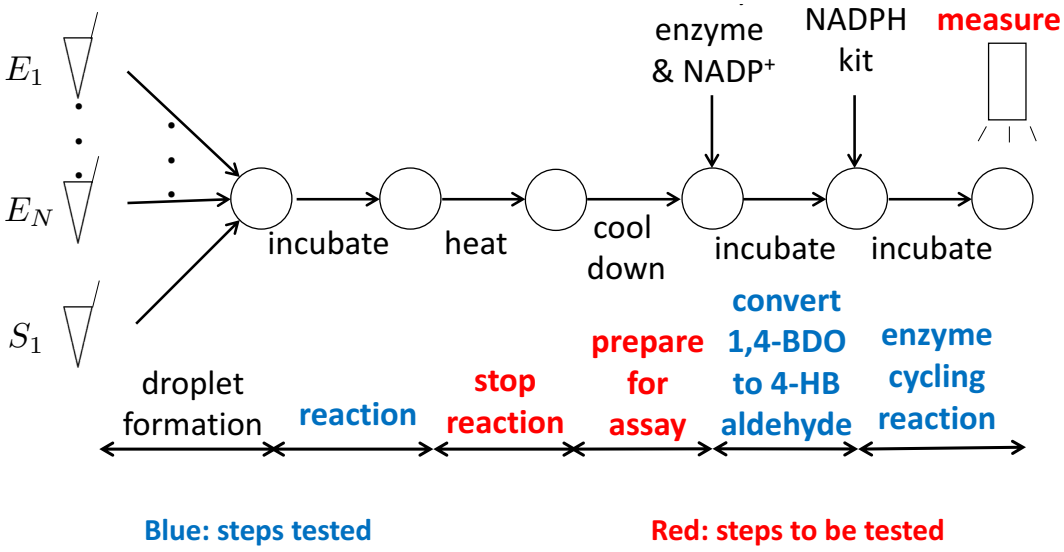


Figure B.3: Overview of experimental plan.

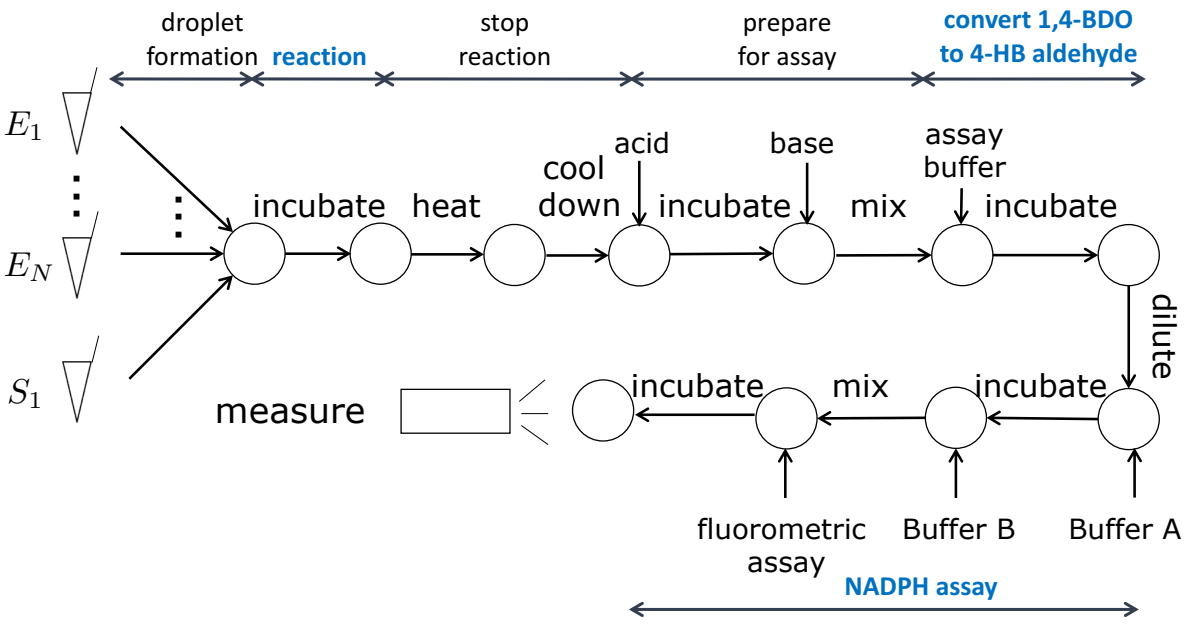


Figure B.4: Detailed schematics for TX-TL in droplets.

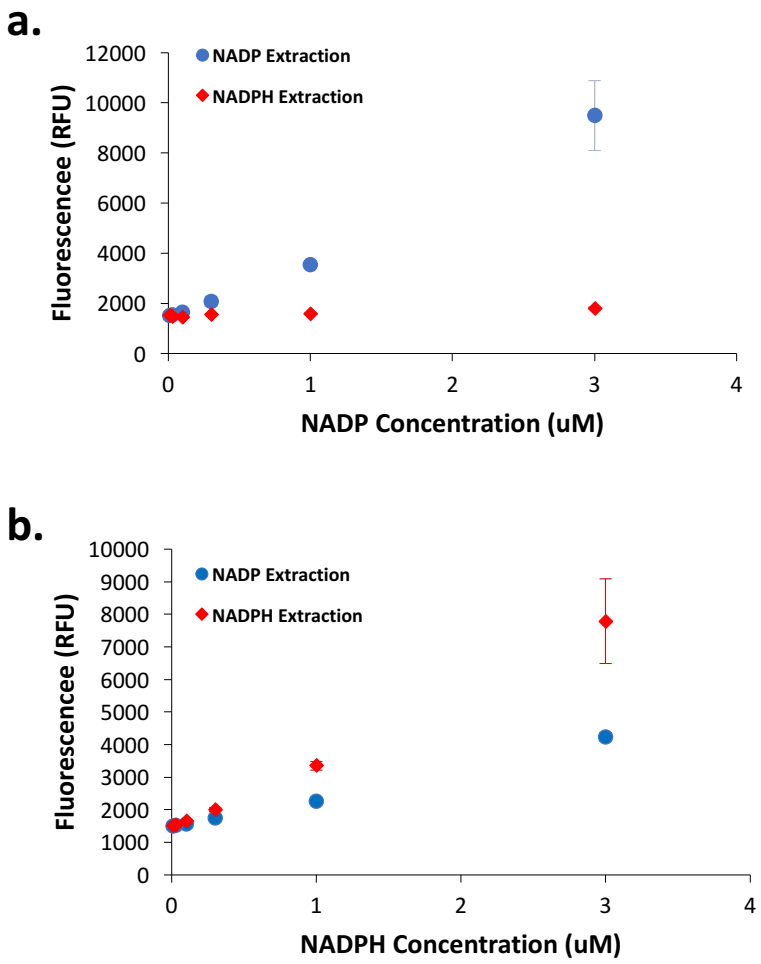


Figure B.5: Testing the robustness of the NADP/NADPH kit a. NADP measured using NADP Extraction and NADPH Extraction b. NADPH measured using NADP Extraction and NADPH Extraction.

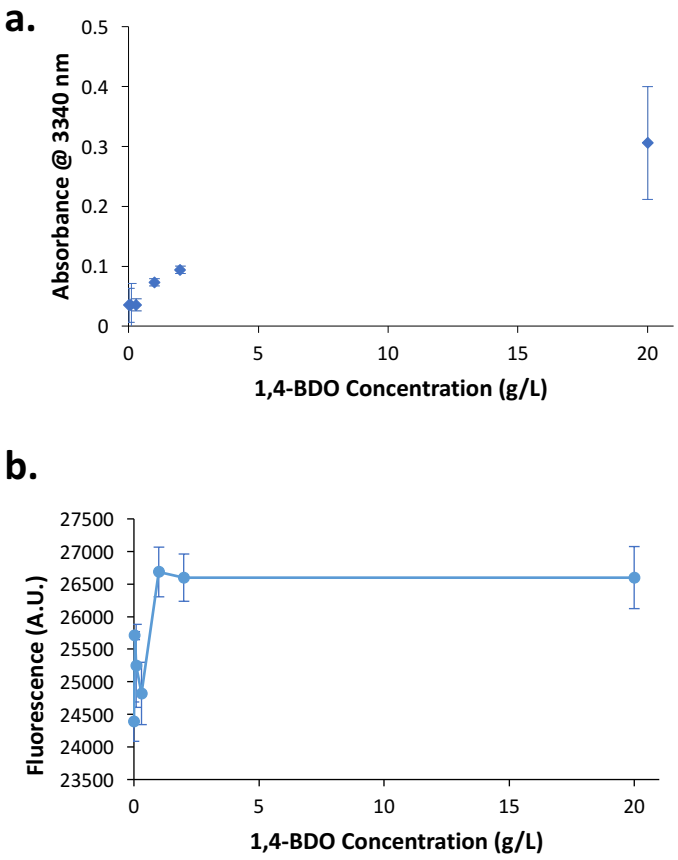


Figure B.6: Results from assay enzymes and NADP/NADPH kit: a. Absorbance of BDO assay, b. Fluorescence of NADPH kit.

BIBLIOGRAPHY

1. Peralta-Yahya, P. P.; Keasling, J. D., Advanced biofuel production in microbes. *Biotechnology Journal* **2010,** *5* (2), 147-162.

2. Lee, J. W.; Na, D.; Park, J. M.; Lee, J.; Choi, S.; Lee, S. Y., Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat Chem Biol* **2012,** *8* (6), 536-546.

3. Jang, Y.-S.; Kim, B.; Shin, J. H.; Choi, Y. J.; Choi, S.; Song, C. W.; Lee, J.; Park, H. G.; Lee, S. Y., Bio-based production of C2–C6 platform chemicals. *Biotechnology and Bioengineering* **2012,** *109* (10), 2437-2459.

4. Woolston, B. M.; Edgar, S.; Stephanopoulos, G., Metabolic Engineering: Past and Future. *Annual Review of Chemical and Biomolecular Engineering* **2013,** *4* (1), 259-288.

5. Orth, J. D.; Thiele, I.; Palsson, B. O., What is flux balance analysis? *Nat Biotech* **2010,** *28* (3), 245-248.

6. Wiechert, W., 13C Metabolic Flux Analysis. *Metabolic Engineering* **2001,** *3* (3), 195-206.

7. Nakamura, C. E.; Whited, G. M., Metabolic engineering for the microbial production of 1,3-propanediol. *Current Opinion in Biotechnology* **2003,** *14* (5), 454-459.

8. Arnold, F. H., Directed evolution: Creating biocatalysts for the future. *Chemical Engineering Science* **1996,** *51* (23), 5091-5102.

9. Coelho, P. S.; Wang, Z. J.; Ener, M. E.; Baril, S. A.; Kannan, A.; Arnold, F. H.; Brustad, E. M., A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat Chem Biol* **2013,** *9* (8), 485-487.

10. Lee, M. E.; Aswani, A.; Han, A. S.; Tomlin, C. J.; Dueber, J. E., Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Research* **2013,** *41* (22), 10668-10678.

11. Kuriyan, J.; Eisenberg, D., The origin of protein interactions and allostery in colocalization. *Nature* **2007,** *450* (7172), 983-990.

12. Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L. J.; Keasling, J. D., Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotech* **2009,** *27* (8), 753-759.

13. Dunlop, M. J.; Dossani, Z. Y.; Szmidt, H. L.; Chu, H. C.; Lee, T. S.; Keasling, J. D.; Hadi, M. Z.; Mukhopadhyay, A., Engineering microbial biofuel tolerance and export using efflux pumps. *Mol Syst Biol* **2011,** *7*.

14. Farmer, W. R.; Liao, J. C., Improving lycopene production in Escherichia coli by engineering metabolic control. *Nat Biotech* **2000,** *18* (5), 533-537.

15. Tang, S.-Y.; Cirino, P. C., Elucidating residue roles in engineered variants of AraC regulatory protein. *Protein Science* **2010,** *19* (2), 291-298.

16. Smanski, M. J.; Zhou, H.; Claesen, J.; Shen, B.; Fischbach, M.; Voigt, C. A., Synthetic biology to access and expand nature's chemical diversity. *Nature reviews. Microbiology* **2016,** *14* (3), 135-149.

17. Mao, X.-M.; Xu, W.; Li, D.; Yin, W.-B.; Chooi, Y.-H.; Li, Y.-Q.; Tang, Y.; Hu, Y., Epigenetic Genome Mining of an Endophytic Fungus Leads to the Pleiotropic Biosynthesis of Natural Products. *Angewandte Chemie International Edition* **2015,** *54* (26), 7592-7596.

18. Srivastava, S., Kotker, J., Hamilton, S., Ruan, P., Tsui, J., Anderson, J. C., Bodik, R., Seshia, S., , Pathway Synthesis uing the Act Ontology. In *Proceedings of the 4th International Workshop on Bio-Design Automation (IWBDA)*, San Francisco, CA, 2012.

19. Swartz, J. R., Transforming biochemical engineering with cell-free biology. *AIChE Journal* **2012,** *58* (1), 5-13.

20. Jewett, M. C.; Swartz, J. R., Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnology and Bioengineering* **2004,** *86* (1), 19-26.

21. Bottomley, W.; Whitfeld, P. R., Cell-Free Transcription and Translation of Total Spinach Chloroplast DNA. *European Journal of Biochemistry* **1979,** *93* (1), 31-39.

22. Roberts, B. E.; Gorecki, M.; Mulligan, R. C.; Danna, K. J.; Rozenblatt, S.; Rich, A., Simian virus 40 DNA directs synthesis of authentic viral polypeptides in a linked transcription-translation cell-free system. *Proceedings of the National Academy of Sciences* **1975,** *72* (5), 1922-1926.

23. Yang, H. L.; Ivashkiv, L.; Chen, H. Z.; Zubay, G.; Cashel, M., Cell-free coupled transcription-translation system for investigation of linear DNA segments. *Proceedings of the National Academy of Sciences* **1980,** *77* (12), 7029-7033.

24. Zawada, J. F.; Yin, G.; Steiner, A. R.; Yang, J.; Naresh, A.; Roy, S. M.; Gold, D. S.; Heinsohn, H. G.; Murray, C. J., Microscale to manufacturing scale-up of cell-free cytokine production—a new approach for shortening protein production development timelines. *Biotechnology and Bioengineering* **2011,** *108* (7), 1570-1578.

25. Fischer, S., Cell Break: How Cell-Free Biology Is Finally Putting the Engineering Back in Bioengineering. *IEEE Pulse* **2016,** *7* (2), 13-16.

26. Noireaux, V.; Bar-Ziv, R.; Libchaber, A., Principles of cell-free genetic circuit assembly. *Proceedings of the National Academy of Sciences* **2003,** *100* (22), 12672-12677.

27. Niederholtmeyer, H.; Sun, Z. Z.; Hori, Y.; Yeung, E.; Verpoorte, A.; Murray, R. M.; Maerkl, S. J., Rapid cell-free forward engineering of novel genetic ring oscillators. *eLife* **2015,** *4*, e09771.

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

29. Takahashi, M. K.; Chappell, J.; Hayes, C. A.; Sun, Z. Z.; Kim, J.; Singhal, V.; Spring, K. J.; Al-Khabouri, S.; Fall, C. P.; Noireaux, V.; Murray, R. M.; Lucks, J. B., Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription–Translation (TX-TL) Systems. *ACS Synthetic Biology* **2015,** *4* (5), 503-515.

30. 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. *Nat Commun* **2014,** *5*.

31. Wang, Y.; Huang, W.; Sathitsuksanoh, N.; Zhu, Z.; Zhang, Y. H. P., Biohydrogenation from Biomass Sugar Mediated by In Vitro Synthetic Enzymatic Pathways. *Chemistry & Biology* **2011,** *18* (3), 372-380.

32. Opgenorth, P. H.; Korman, T. P.; Bowie, J. U., A synthetic biochemistry module for production of bio-based chemicals from glucose. *Nat Chem Biol* **2016,** *12* (6), 393-395.

33. Sun, Z. Z.; Hayes, C. A.; Shin, J.; Caschera, F.; Murray, R. M.; Noireaux, V., Protocols for Implementing an Escherichia coli Based TX-TL Cell-Free Expression System for Synthetic Biology. **2013,** (79), e50762.

34. Li, G.-W.; Burkhardt, D.; Gross, C.; Weissman, J. S., Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **2014,** *157* (3), 624-635.

35. Tan, S. Z.; Manchester, S.; Prather, K. L. J., Controlling Central Carbon Metabolism for Improved Pathway Yields in Saccharomyces cerevisiae. *ACS Synthetic Biology* **2016,** *5* (2), 116-124.

36. Bogorad, I. W.; Lin, T.-S.; Liao, J. C., Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature* **2013,** *502* (7473), 693-697.

37. de Jong, H.; Geiselmann, J.; Ropers, D., Resource Reallocation in Bacteria by Reengineering the Gene Expression Machinery. *Trends in Microbiology* **2017,** *25* (6), 480-493.

38. Scott, M.; Gunderson, C. W.; Mateescu, E. M.; Zhang, Z.; Hwa, T., Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science* **2010,** *330* (6007), 1099-1102.

39. Glick, B. R., Metabolic load and heterologous gene expression. *Biotechnology Advances* **1995,** *13* (2), 247-261.

40. Drews, J., Drug Discovery: A Historical Perspective. *Science* **2000,** *287* (5460), 1960.

41. Schreiber, S. L., Target-Oriented and Diversity-Oriented Organic Synthesis in Drug Discovery. *Science* **2000,** *287* (5460), 1964.

42. Trosset, J.-Y.; Carbonell, P., Synthetic biology for pharmaceutical drug discovery. *Drug Design, Development and Therapy* **2015,** *9*, 6285-6302.

43. Koehn, F. E.; Carter, G. T., The evolving role of natural products in drug discovery. *Nat Rev Drug Discov* **2005,** *4* (3), 206-220.

44. Dewick, P. M., *Medicinal Natural Products: A Biosynthetic Approach*. Wiley: 2002.

45. Newman, D. J.; Cragg, G. M., Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *Journal of Natural Products* **2012,** *75* (3), 311-335.

46. Ro, D.-K.; Paradise, E. M.; Ouellet, M.; Fisher, K. J.; Newman, K. L.; Ndungu, J. M.; Ho, K. A.; Eachus, R. A.; Ham, T. S.; Kirby, J.; Chang, M. C. Y.; Withers, S. T.; Shiba, Y.; Sarpong, R.; Keasling, J. D., Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006,** *440* (7086), 940-943.

47. Tu, Y., The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nat Med* **2011,** *17* (10), 1217-1220.

48. Socha, A. M.; Tan, N. Y.; LaPlante, K. L.; Sello, J. K., Diversity-oriented synthesis of cyclic acyldepsipeptides leads to the discovery of a potent antibacterial agent. *Bioorganic & Medicinal Chemistry* **2010,** *18* (20), 7193-7202.

49. Davis, M. R.; Dougherty, D. A., Cation-π Interactions: Computational Analyses of the Aromatic Box Motif and the Fluorination Strategy for Experimental Evaluation. *Physical chemistry chemical physics: PCCP* **2015,** *17* (43), 29262-29270.

50. Fleming, F. F.; Yao, L.; Ravikumar, P. C.; Funk, L.; Shook, B. C., Nitrile-Containing Pharmaceuticals: Efficacious Roles of the Nitrile Pharmacophore. *Journal of medicinal chemistry* **2010,** *53* (22), 7902-7917.

51. Gillis, E. P.; Eastman, K. J.; Hill, M. D.; Donnelly, D. J.; Meanwell, N. A., Applications of Fluorine in Medicinal Chemistry. *Journal of Medicinal Chemistry* **2015,** *58* (21), 8315-8359.

52. Schönherr, H.; Cernak, T., Profound Methyl Effects in Drug Discovery and a Call for New C-H Methylation Reactions. *Angewandte Chemie International Edition* **2013,** *52* (47), 12256-12267.

53. Arnold, F. H., The nature of chemical innovation: new enzymes by evolution. *Quarterly Reviews of Biophysics* **2015,** *48* (4), 404-410.

54. Maier, M. E., Design and synthesis of analogues of natural products. *Organic & Biomolecular Chemistry* **2015,** *13* (19), 5302-5343.

55. Wang, Z. J.; Renata, H.; Peck, N. E.; Farwell, C. C.; Coelho, P. S.; Arnold, F. H., Improved Cyclopropanation Activity of Histidine-Ligated Cytochrome P450 Enables the Enantioselective Formal Synthesis of Levomilnacipran. *Angewandte Chemie International Edition* **2014,** *53* (26), 6810-6813.

56. Murciano-Calles, J.; Romney, D. K.; Brinkmann-Chen, S.; Buller, A. R.; Arnold, F. H., A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the Transfer of Mutations that Mimic Allosteric Activation. *Angewandte Chemie International Edition* **2016,** *55* (38), 11577-11581.

57. Alzari, P. M.; Berglund, H.; Berrow, N. S.; Blagova, E.; Busso, D.; Cambillau, C.; Campanacci, V.; Christodoulou, E.; Eiler, S.; Fogg, M. J.; Folkers, G.; Geerlof, A.; Hart, D.; Haouz, A.; Herman, M. D.; Macieira, S.; Nordlund, P.; Perrakis, A.; Quevillon-Cheruel, S.; Tarandeau, F.; van Tilbeurgh, H.; Unger, T.; Luna-Vargas, M. P. A.; Velarde, M.; Willmanns, M.; Owens, R. J., Implementation of semi-automated cloning and prokaryotic expression screening: the impact of SPINE. *Acta Crystallographica Section D* **2006,** *62* (10), 1103-1113.

58. Alper, H.; Miyaoku, K.; Stephanopoulos, G., Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nat Biotech* **2005,** *23* (5), 612-616.

59. Hong, K.-K.; Nielsen, J., Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries. *Cellular and Molecular Life Sciences* **2012,** *69* (16), 2671-2690.

60. 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* **2014,** *3* (6), 387-397.

61. Ng, P. P.; Jia, M.; Patel, K. G.; Brody, J. D.; Swartz, J. R.; Levy, S.; Levy, R., A vaccine directed to B cells and produced by cell-free protein synthesis generates potent antilymphoma immunity. *Proceedings of the National Academy of Sciences* **2012,** *109* (36), 14526-14531.

62. Pardee, K.; Slomovic, S.; Nguyen, Peter Q.; Lee, Jeong W.; Donghia, N.; Burrill, D.; Ferrante, T.; McSorley, Fern R.; Furuta, Y.; Vernet, A.; Lewandowski, M.; Boddy, Christopher N.; Joshi, Neel S.; Collins, James J., Portable, On-Demand Biomolecular Manufacturing. *Cell* **2016,** *167* (1), 248-259.e12.

63. Kay, J. E.; Jewett, M. C., Lysate of engineered Escherichia coli supports high-level conversion of glucose to 2,3-butanediol. *Metabolic Engineering* **2015,** *32*, 133-142.

64. Karim, A. S.; Jewett, M. C., A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. *Metabolic Engineering* **2016,** *36*, 116-126.

65. Shin, J.; Noireaux, V., An E. coli Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. *ACS Synthetic Biology* **2011,** *1* (1), 29-41.

66. Caschera, F.; Noireaux, V., A cost-effective polyphosphate-based metabolism fuels an all E. coli cell-free expression system. *Metabolic Engineering* **2015,** *27*, 29-37.

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

68. 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. *Molecular Systems Biology* **2008,** *4*, 220-220.

69. 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* **2004,** *87* (4), 465-471.

70. Culler, S., An integrated bioengineering platform that harnesses the synthetic biology toolbox is the key to an economically viable commercial bioprocess. *Chemical Engineering Progress* 2016, pp 42-51.

71. Du, J.; Yuan, Y.; Si, T.; Lian, J.; Zhao, H., Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Research* **2012,** *40* (18), e142.

72. Lee, M. E.; Aswani, A.; Han, A. S.; Tomlin, C. J.; Dueber, J. E., Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Research* **2013**.

73. Coussement, P.; Maertens, J.; Beauprez, J.; Van Bellegem, W.; De Mey, M., One step DNA assembly for combinatorial metabolic engineering. *Metabolic Engineering* **2014,** *23*, 70-77.

74. 1,4-Butanediol Market by Technology (Reppe process, Davy process, Butadiene process, Propylene oxide process and others), and by Application (THF, PBT, GBL, PU, and Others) - Global Trends & Forecasts to 2019. http://www.marketsandmarkets.com/Market-Reports/1-4-butanediol-market-685.html (accessed February 17, 2017).

75. Novamont opens world's first commercial plant for bio-production of a major intermediate chemical. https://www.genomatica.com/ (accessed March 20, 2017).

76. Burgard, A.; Burk, M. J.; Osterhout, R.; Van Dien, S.; Yim, H., Development of a commercial scale process for production of 1,4-butanediol from sugar. *Current Opinion in Biotechnology* **2016,** *42*, 118-125.

77. Mutalik, V. K.; Guimaraes, J. C.; Cambray, G.; Lam, C.; Christoffersen, M. J.; Mai, Q.-A.; Tran, A. B.; Paull, M.; Keasling, J. D.; Arkin, A. P.; Endy, D., Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Meth* **2013,** *10* (4), 354-360.

78. Yim, H.; Haselbeck, R.; Niu, W.; Pujol-Baxley, C.; Burgard, A.; Boldt, J.; Khandurina, J.; Trawick, J. D.; Osterhout, R. E.; Stephen, R.; Estadilla, J.; Teisan, S.; Schreyer, H. B.; Andrae, S.; Yang, T. H.; Lee, S. Y.; Burk, M. J.; Van Dien, S., Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nat Chem Biol* **2011,** *7* (7), 445-452.

79. Barton, N. R.; Burgard, A. P.; Burk, M. J.; Crater, J. S.; Osterhout, R. E.; Pharkya, P.; Steer, B. A.; Sun, J.; Trawick, J. D.; Van Dien, S. J.; Yang, T. H.; Yim, H., An integrated biotechnology platform for developing sustainable chemical processes. *J Ind Microbiol Biotechnol* **2015,** *42* (3), 349-360.

80. Tuza, Z. A., Singhal, V., Kim, J., Murray, R. M., An In Silico Modeling Toolbox for Rapid Prototyping of Circuits in a Biomolecular "Breadboard" System. In *52nd IEEE Conference on Decision and Control*, Florence, Italy, 2013.

81. 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* **2014,** *3* (6), 416-425.

82. Gilda, J. E.; Gomes, A. V., Stain Free Total Protein Staining is a Superior Loading Control to β-Actin for Western Blots. *Analytical Biochemistry* **2013,** *440* (2), 10.1016/j.ab.2013.05.027.

83. Kao, P., Re: Ladder protein cocentration. Wu, Y. Y., Ed. 2017.

84. Yeung, E.; Dy, A. J.; Martin, K. B.; Ng, A. H.; Del Vecchio, D.; Beck, J. L.; Collins, J. J.; Murray, R. M., The Effect of Compositional Context on Synthetic Gene Networks. *bioRxiv* **2016**.

85. Lutz, R.; Bujard, H., Independent and Tight Regulation of Transcriptional Units in Escherichia Coli Via the LacR/O, the TetR/O and AraC/I1-I2 Regulatory Elements. *Nucleic Acids Research* **1997,** *25* (6), 1203-1210.

86. Kwon, Y.-C.; Jewett, M. C., High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Scientific Reports* **2015,** *5*, 8663.

87. Fraser, M. E.; James, M. N. G.; Bridger, W. A.; Wolodko, W. T., A detailed structural description of Escherichia coli succinyl-CoA synthetase. *Journal of Molecular Biology* **1999,** *288* (3), 501.

88. 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. *Molecular Systems Biology* **2008,** *4* (1), n/a-n/a.

89. Chen, Y.-J.; Liu, P.; Nielsen, A. A. K.; Brophy, J. A. N.; Clancy, K.; Peterson, T.; Voigt, C. A., Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat Meth* **2013,** *10* (7), 659-664.

90. Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J., DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research* **2006,** *34* (Database issue), D668-D672.

91. Kan, S. B. J.; Lewis, R. D.; Chen, K.; Arnold, F. H., Directed evolution of cytochrome c for carbon–silicon bond formation: Bringing silicon to life. *Science* **2016,** *354*, 1048-1051.

92. Hernandez, K. E.; Renata, H.; Lewis, R. D.; Kan, S. B. J.; Zhang, C.; Forte, J.; Rozzell, D.; McIntosh, J. A.; Arnold, F. H., Highly Stereoselective Biocatalytic Synthesis of Key Cyclopropane Intermediate to Ticagrelor. *ACS Catalysis* **2016,** *6* (11), 7810-7813.

93. Carney, D. W.; Schmitz, K. R.; Truong, J. V.; Sauer, R. T.; Sello, J. K., Restriction of the Conformational Dynamics of the Cyclic Acyldepsipeptide Antibiotics Improves Their Antibacterial Activity. *Journal of the American Chemical Society* **2014,** *136* (5), 1922-1929.

94. DeMoss, R. D.; Evans, N. R., INCORPORATION OF C(14)-LABELED SUBSTRATES INTO VIOLACEIN. *Journal of Bacteriology* **1960,** *79* (5), 729-733.

95. Queiroz, K. C. S.; Milani, R.; Ruela-de-Sousa, R. R.; Fuhler, G. M.; Justo, G. Z.; Zambuzzi, W. F.; Duran, N.; Diks, S. H.; Spek, C. A.; Ferreira, C. V.; Peppelenbosch, M. P., Violacein Induces Death of Resistant Leukaemia Cells via Kinome Reprogramming, Endoplasmic Reticulum Stress and Golgi Apparatus Collapse. *PLOS ONE* **2012,** *7* (10), e45362.

96. de Carvalho, D. D.; Costa, F. T. M.; Duran, N.; Haun, M., Cytotoxic activity of violacein in human colon cancer cells. *Toxicology in Vitro* **2006,** *20* (8), 1514-1521.

97. Lopes, S. C. P.; Blanco, Y. C.; Justo, G. Z.; Nogueira, P. A.; Rodrigues, F. L. S.; Goelnitz, U.; Wunderlich, G.; Facchini, G.; Brocchi, M.; Duran, N.; Costa, F. T. M., Violacein Extracted from Chromobacterium violaceum Inhibits Plasmodium Growth In Vitro and In Vivo. *Antimicrobial Agents and Chemotherapy* **2009,** *53* (5), 2149-2152.

98. Choi, S. Y.; Yoon, K.-h.; Lee, J. I.; Mitchell, R. J., Violacein: Properties and Production of a Versatile Bacterial Pigment. *BioMed Research International* **2015,** *2015*, 8.

99. Durán, N.; Justo, G. Z.; Ferreira, C. V.; Melo, P. S.; Cordi, L.; Martins, D., Violacein: properties and biological activities. *Biotechnology and Applied Biochemistry* **2007,** *048* (3), 127-133.

100. Durán, M.; Faljoni-Alario, A.; Durán, N., Chromobacterium violaceum and its important metabolites — review. *Folia Microbiol* **2010,** *55* (6), 535-547.

101. Balibar, C. J.; Walsh, C. T., In Vitro Biosynthesis of Violacein from l-Tryptophan by the Enzymes VioA−E from Chromobacterium violaceum. *Biochemistry* **2006,** *45* (51), 15444-15457.

102. Nguyen, P. H. B.; Wu, Y.; Guo, S.; Murray, R. M., Design Space Exploration of the Violacein Pathway in Escherichia coli Based Transcription Translation Cell-Free System (TX-TL). *bioRxiv* **2015**.

103. 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* **2016,** *5* (4), 344-355.

104. Howard-Jones, A. R.; Walsh, C. T., Enzymatic Generation of the Chromopyrrolic Acid Scaffold of Rebeccamycin by the Tandem Action of RebO and RebD. *Biochemistry* **2005,** *44* (48), 15652-15663.

105. Trott, O.; Olson, A. J., AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **2010,** *31* (2), 455-461.

106. Ryan, K. S.; Balibar, C. J.; Turo, K. E.; Walsh, C. T.; Drennan, C. L., The violacein biosynthetic enzyme vioe shares a fold with lipoprotein transporter proteins. *Journal of Biological Chemistry* **2008,** *283*, 6467-6475.

107. Salazar, O.; Sun, L., Evaluating a Screen and Analysis of Mutant Libraries. In *Directed Enzyme Evolution: Screening and Selection Methods*, Arnold, F. H.; Georgiou, G., Eds. Springer Science & Business Media: 2003; Vol. 230, pp 85-97.

108. Schuttelkopf, A. W.; van Aalten, D. M. F., PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D* **2004,** *60* (8), 1355-1363.

109. Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J., Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protocols* **2016,** *11* (5), 905-919.

110. Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z.-G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P., Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synthetic Biology* **2012,** *2* (2), 83-92.

111. Martin, R. W.; Majewska, N. I.; Chen, C. X.; Albanetti, T. E.; Jimenez, R. B. C.; Schmelzer, A. E.; Jewett, M. C.; Roy, V., Development of a CHO-Based Cell-Free Platform for Synthesis of Active Monoclonal Antibodies. *ACS Synthetic Biology* **2017,** *6* (7), 1370-1379.

112. Li, J.; Wang, H.; Kwon, Y.-C.; Jewett, M. C., Establishing a high yielding streptomyces-based cell-free protein synthesis system. *Biotechnology and Bioengineering* **2017,** *114* (6), 1343-1353.

113. Zhu, Z.; Wang, Y.; Minteer, S. D.; Percival Zhang, Y. H., Maltodextrin-powered enzymatic fuel cell through a non-natural enzymatic pathway. *Journal of Power Sources* **2011,** *196* (18), 7505-7509.

114. Guo, S.; Yeung, E.; Nilgiriwala, K. S.; Del Vecchio, D.; Murray, R. M. In *Implementation And Simulation Of Phosphorylation-Based Insulator In Transcription-Translation Platform*, Winter Q-Bio Conference: 2014.

115. Yeung, E.; Ng, A.; Kim, J.; Sun, Z. Z.; Murray, R. M. In *Modeling the effects of compositional context on promoter activity in an E. coli extract based transcription-translation system*, Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on, IEEE: 2014; pp 5405-5412.

116. Murray, R. M.; Singhal, V., Cell-free biomolecular breadboards and related methods and arrangements. Google Patents: 2016.

117. Yang, C.-H.; Li, Y.-H., Chromobacterium violaceum infection: A clinical review of an important but neglected infection. *Journal of the Chinese Medical Association* **2011,** *74* (10), 435-441.

118. Kumar, M. R., Chromobacterium violaceum: A rare bacterium isolated from a wound over the scalp. *International Journal of Applied and Basic Medical Research* **2012,** *2* (1), 70-72.

119. Magee, R.; Kosaric, N., The microbial production of 2,3-butanediol. *Adv Appl Microbial* **1987,** *32*, 89 - 161.

120. Zeng, A.-P.; Sabra, W., Microbial production of diols as platform chemicals: Recent progresses. *Current Opinion in Biotechnology* **2011,** *22* (6), 749-757.

121. Ma, C.; Wang, A.; Qin, J.; Li, L.; Ai, X.; Jiang, T.; Tang, H.; Xu, P., Enhanced 2,3-butanediol production by Klebsiella pneumoniae SDM. *Appl Microbiol Biotechnol* **2009,** *82*, 49 - 57.

122. Zhang, L., Sun, J., Hao, Y., Zhu, J., Chu, J., Wei, D., Shen, Y., Microbial production of 2,3-butanediol by a surfactant (serrawettin)-deficient mutant of Serratia marcescens H30. *J Ind Microbiol Biotechnol* **2010,** *37* (8), 857-862.

123. O'Sullivan, S. M.; Condon, S.; Cogan, T. M.; Sheehan, D., Purification and characterisation of acetolactate decarboxylase from Leuconostoc lactis NCW1. *FEMS Microbiology Letters* **2001,** *194* (2), 245-249.

124. Yan, Y.; Lee, C.-C.; Liao, J. C., Enantioselective synthesis of pure (R,R)-2,3-butanediol in Escherichia coli with stereospecific secondary alcohol dehydrogenases. *Organic & Biomolecular Chemistry* **2009,** *7* (19), 3914-3917.

125. Atsumi, S.; Li, Z.; Liao, J. C., Acetolactate synthase from Bacillus subtilis serves as a 2-ketoisovalerate decarboxylase for isobutanol biosynthesis in Escherichia coli. *Applied and environmental microbiology* **2009,** *75* (19), 6306-6311.

126. Siegal-Gaskins, D.; Tuza, Z. A.; Kim, J.; Noireaux, V.; Murray, R. M., Resource usage and gene circuit performance characterization in a cell-free "breadboard". *bioRxiv* **2014**.

127. Quax, Tessa E. F.; Wolf, Yuri I.; Koehorst, Jasper J.; Wurtzel, O.; van der Oost, R.; Ran, W.; Blombach, F.; Makarova, Kira S.; Brouns, Stan J. J.; Forster, Anthony C.; Wagner, E. Gerhart H.; Sorek, R.; Koonin, Eugene V.; van der Oost, J., Differential Translation Tunes Uneven Production of Operon-Encoded Proteins. *Cell Reports* **2013,** *4* (5), 938-944.

128. Karzbrun, E.; Shin, J.; Bar-Ziv, R. H.; Noireaux, V., Coarse-Grained Dynamics of Protein Synthesis in a Cell-Free System. *Physical Review Letters* **2011,** *106* (4), 048104.

129. O'Sullivan, S. M., Condon, S., Cogan, T. M., Sheehan, D., Purification and characterisation of acetolactate decarboxylase from Leuconostoc lactis NCW1. *FEMS Microbiol. Lett.* **2001,** *194* (2), 245-9.

130. Nielsen, D.; Yoon, S.; Yuan, C.; Prather, K., Metabolic engineering of acetoin and meso-2,3-butanediol biosynthesis in E. coli. *Biotechnol J* **2010,** *5*, 274 - 284.

131. Tran, T. M.; Lan, F.; Thompson, C. S.; Abate, A. R., From tubes to drops: droplet-based microfluidics for ultrahigh-throughput biology. *Journal of Physics D: Applied Physics* **2013,** *46* (11), 114004.