Investigating and Developing Strategies for the Expression of Biosynthetic Genes *In Vitro* and *In Vivo*

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Tien Thuy Tran
May 2024

DEDICATION

I dedicate this dissertation to my family.

To my parents, Tran Ngoc Canh and Nguyen Thi Thu Thuy, who sacrificed so much and are my best support system. Dad, you have had so much faith in me ever since I was a little girl. What motivated me to be who I am today was the poem you wrote for me while you were being drafted into the military, wishing me to have a better life than yours and have opportunities to pursue my dreams. Mom, you are the reason I fell in love with math and science. You are and will be forever the wisest person in my life, who will always be the first one I look to for comfort and support.

To my puppy Xena, I adopted you in the first semester of my PhD journey. You have brightened my quarantine days and emotionally supported me through many mental breakdown days.

And lastly, to my best friend, my wonderful husband Justin Sword, thank you for staying by my side and supporting me during my hardest time. You have always been my deep source of emotional support. We overcame the pandemic COVID-19 and grad school together. There is a life-long to go.

I would not make it through this important milestone without all of your love and support. Love you all so much.

ACKNOWLEDGEMENTS

Graduate school has been a journey that I would not have been able to accomplish without endless help and support from those around me. I would like to thank everyone who has been along with me on my Ph. D journey. First and foremost, my deepest gratitude goes to my mentor Prof. Constance B. Bailey for her limitless support and guidance along every stage of my journey. I sincerely appreciate her time patiently providing feedback and advice to improve my research and writing skills. Words cannot express my appreciation and gratitude for such a mentor who tremendously helped me grow not only as a scientist but also as a person.

I am truly grateful for the opportunity to collaborate with Dr. Mitchel J. Doktycz and his graduate student Dr. Jaime Lorenzo Naval Dinglasan at Oak Ridge National Laboratory. I extremely appreciate Dr. Doktycz advice and his support while working with him, especially during COVID and after my advisor moved to another continent. Enzo is the best lab partner I could ever ask for. Thanks for being patient with me and guiding me through the cell-free project, we had two very nice papers together. I am so thankful I get to collaborate with him, we had such an unforgettable time at the SIMB conference in San Diego.

I would like to thank Prof. Tessa R. Calhoun for countless meetings in her office. She is my co-advisor who has guided my intellectual development to be an independent thinker. Thanks for welcoming me to your group and taking over the mentor role after Dr. Bailey moved.

I greatly appreciate Prof. Michael D. Best for agreeing to serve on my Ph.D. committee. Thanks for your time and valuable comments throughout my journey.

I would like to thank Prof. Michael D. Burkart from the University of California, San Diego, for a useful discussion about the Rosetta plasmid. I finally published that paper using the plasmid DNA from his laboratory.

I also would like to thank the members of the Bailey and the Calhoun Research Group. Especially Marea Blake, who helped me so much with my drafts and practice talks for both of my qualifying exams.

Most importantly, I thank my husband Justin Sword, who was by my side and motivated me on my lowest days. Thanks for sitting through all the science talks, improving my presentations, and motivating me to be who I am today.

ABSTRACT

Natural products have historically served as a rich source for a wide range of useful applications such as pesticides, veterinary agents, therapeutics, and bioproducts. To discover new natural products, manipulate them for analog generation, and harness the potential of these bioactive compounds for synthetic biology, it is necessary to develop robust methods for the expression of biosynthetic genes. A broad range of clinically useful natural products originate from actinomycetes, especially those from the genus *Streptomyces,* which have been recognized as one of the predominant sources of microbial bioactive natural products. Actinobacteria are known for their large genomes, guanine-cytosine rich, and complex secondary metabolism. Some of these secondary metabolite pathways are composed of multienzyme proteins termed "megasynthases" which produce a wealth of clinically important natural product compounds (e.g., penicillin, daptomycin, and vancomycin antibiotics). The three projects in this dissertation investigate and develop new strategies to express multienzymes originating from actinomycetes *in vivo* and *in vitro*.

Often, the native or engineered pathways must be moved into a suitable surrogate, especially when the native host organism is not genetically tractable. Heterologous hosts must be genetically manipulatable and generally contain both sufficient fluxes of metabolic precursor as well as an appropriate environment for protein folding and expression. The first project of this dissertation reveals the effects of refactoring biosynthetic megasynthases using *Escherichia coli* as a heterologous host to express a non-ribosomal peptide synthase originating from *Streptomyces*.

Microbial metabolic engineering has focused on creating "cell factories" that can synthesize valuable metabolites from readily available substrates. Cell-free synthetic biology is emerging as an important complementary approach because it is highly desirable to express protein on a more rapid timescale and does not rely upon the genetic tractability of a strain thus improving the throughput of design-build-test-learn (DBTL) cycles. The second project focuses on investigating and optimizing lysate-based expression for megasynthase proteins.

The development of cell-free systems can then be used for prototyping these complex pathways to accelerate efforts towards engineered biosynthesis of these pathways. Taking advantage of that, the third project uses a lysate-base cell-free system platform to profile expression strategies for multienzymes expression.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

CHAPTER 1

INTRODUCTION

Parts of this chapter were originally published by Tien T Sword, Ghaeath S.K. Abbas, and Constance B Bailey:

Sword TT, Abbas GSK, Bailey CB. Cell-Free Protein Synthesis for Nonribosomal Peptide Synthetic Biology. Frontiers in Natural Products. 2024 Jan18;

1.1 Importance of Natural Products in Human Life

Natural products are complex secondary metabolites that produce a wealth of bioactive compounds. Historically, natural product discovery and development have been fields of interest for a range of applications including biomedicine, (1) agrochemicals, (2–5) biocatalyst discovery, (6–8) and byproduct formation. (9–11) Of these applications, the most historically significant one has been as an important source for drug discovery. (1,12) Derived from all kingdoms of life, natural products continue to account for a substantial portion of approved therapeutics. In fact, about 75% of today's antibiotics are desired from natural product chemical scaffolds. (13,14) From 1981 to 2019, 32% of all small-molecule drugs approved were natural products and their derivatives. (15–17) Moreover, among 136 small-molecule anticancer drugs found from 1940 to 2014, 68% were based on natural product scaffolds. (18)

Bioactive natural products are traditionally extracted from their native producers directly. Since the beginning of the golden age of antibiotics (between 1940 and the early 1960s), extensive efforts have been made to isolate various highly complex natural products by using bioactivity-guided screening. (19) In fact, most of the natural product-derived drugs that we are using today are the result of these traditional extraction efforts. (20) However, it can be challenging to harvest natural products from these native hosts due to the low productivity or no productivity in a laboratory context (often called "cryptic" or "silent" biosynthetic gene clusters). Additionally, the structures of natural products are often too complex to make total chemical synthesis efficient, practical, and feasible. Thus, due to their numerous applications, it is necessary to get access to these valuable pharmaceutical natural products and develop tools to manipulate them to leverage for drug discovery efforts.

1.2. Polyketide Synthase (PKS) and Non-Ribosomal Peptide Synthetase (NRPS)

Natural products are complex secondary metabolites that produce a wealth of clinically important natural product compounds (e.g., penicillin, daptomycin, and vancomycin antibiotics). (21,22) Some of these secondary metabolite pathways are composed of multienzyme proteins termed "megasynthases". Megasynthases catalyze the serial condensation followed by subsequent processing of precursor units, consisting of two major classes: polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) gene clusters encode for the production of polyketides and non-ribosomal peptides, respectively.

Polyketides are made from the repetitive addition of malonyl-CoA and derivatives, whereas non-ribosomal peptides are biosynthesized by sequential condensation of amino

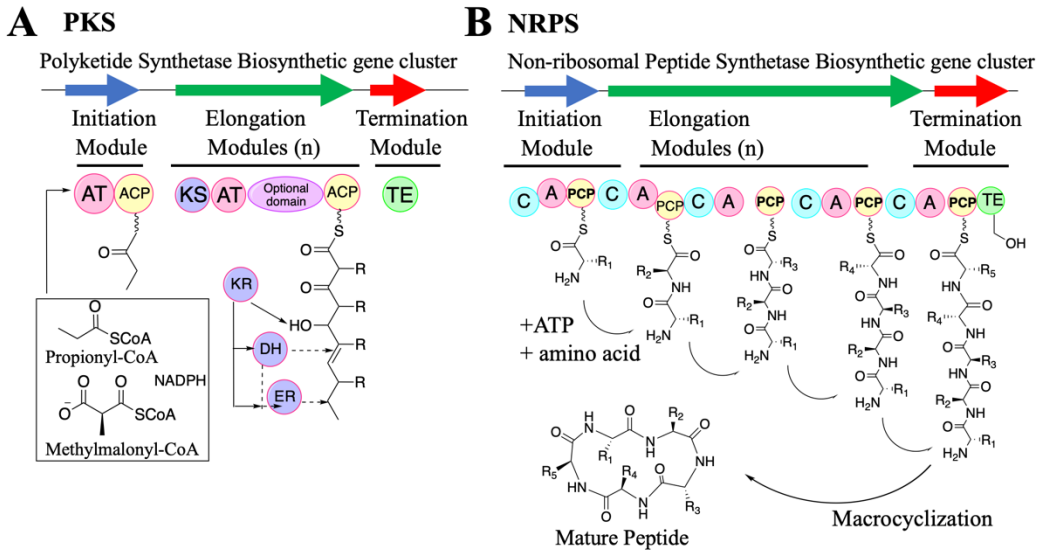


Figure 1.1. Figure Biosynthetic logic of PKS and NRPS.
A. PKS biosynthetic logic. B. NRPS biosynthetic logic.

acid monomers. PKSs and NRPSs are comprised of multi-module proteins, which are sets of distinct active sites for catalyzing each chain elongation and condensation step. Each module in a PKS and NRPS consists of core domains and variables of optional domains responsible for modification of the ketide/peptide backbone. (23–25) PKSs are classified into three main types: noniterative type I PKS, iterative type II PKS, and acyl carrier protein-independent type III PKS. While type I and type II PKSs use acyl carrier protein (ACP) to activate the acyl CoA substrates, type III PKS's act directly on the acyl CoA substrate because it is independent of ACP. (26) A typical PKS module at a minimum consists of three core domains, an acyltransferase (AT) domain, ketosynthase (KS) domain, and an ACP domain. During a single polyketide extension cycle in the PKS system, the KS domain catalyzes a Claisen decarboxylative condensation between an extender unit thioesterified and a starter acyl group to the ACP domain via the action of an AT domain. Other common optional domains of PKS are ketoreductase (KR) domain, dehydratase (DH) domain, and enoylreductase (ER) domain. Finally, located at the C-terminus of the last elongation module, a thioesterase (TE) domain, catalyzes the termination of PK biosynthesis to form a macrolactone natural product. (27)(28,29)

On another hand, the non-ribosomal peptide synthetases are modular enzymes that catalyze the synthesis of peptide products by the formation of peptide bonds. Like PKS, for NRPS, a module is also a set of domains responsible for one cycle of the chain extension in a colinear fashion. A typical NRPS module contains three core domains: an adenylation (A) domain that recognizes and activates an amino acid monomer as an adenylate followed by acyl transfer to a peptidyl carrier protein (PCP) domain, which is often known as thiolation (T) domain along with optional domains such as an epimerase (E) domain or an oxidase (Ox) domain. The activated amino acid is loaded by the PCP domain on a 4′-phosphopantetheine (4′-Ppant) arm, binding it for peptide bond formation with an amino acid on the subsequent module. A peptide bond between two acylated PCP domains during a single peptide extension cycle is formed by the condensation (C) domain. Chain release typically occurs by hydrolysis or macrocyclization via a thioesterase (TE) (**Figure 1.1)** (20,22,30) with other termination mechanisms also sometimes occurring such as reduction to the primary alcohol via a reductase. (31,32) NRPSs must be activated by post-translationally modifying the T domains with a phosphopantetheinyltransferase which provides a tether for the elongating peptide. Due to this colinear relationship between domain organization and metabolite structure, there has been extensive research focused on engineering pipelines of chimeric PKS/NRPS pathways including altering domain selectivity motifs (33,34) or domain swapping. (35–38) For instance, existing antibiotic chemical scaffolds can be modified via engineered biosynthesis to generate analogs that have more targeted or improved activities and/or combat antibiotic resistance.

1.3. Heterologous Expression of Biosynthetic Megasynthase

In addition to bioengineering applications, strategies for heterologous expression are necessary to discover new natural product scaffolds and to access adequate amounts for testing of bioactivity. The discovery of novel natural products is hampered by low productivity or no productivity in a laboratory context (often called “cryptic” or “silent” biosynthetic gene clusters). (39) Because not all of these bioactivities arise from organisms that are tractable to grow in the lab (40) or produce natural products in high abundance, and because the structural complexity of peptide natural products requires chemical synthesis that is often non-trivial, heterologous expression remains a valuable option for characterization and scaleup. (41) One of the most prolific sources of clinically useful natural products is Actinobacteria, which is an order of filamentous bacteria found in diverse ecosystems around the world. (42) Actinobacteria have been recognized as one of the predominant sources of microbial bioactive natural products, particularly the *Streptomyces* genus, which have been sources of chemotherapeutics, (43,44) immunosuppressants, (45–48) antibiotics, (49–52) and anti-parasitic for veterinary agents, (53) pesticides, (54) and pharmaceuticals. (55,56) Actinobacteria are Gram-positive and known for their complex secondary metabolism with rich guanine-cytosine (GC) content. With regard to the morphology (mycelial clumping), growth properties (slow growth), and the fact that most reproducible genetic manipulation relies upon intergenic conjugation, (57,58) *Streptomyces* tend to be less tractable hosts for many applications, such as generating large quantities of protein expression for protein characterization. For these reasons, dedicated campaigns have been pursued to generate complex natural products in more tractable heterologous hosts. (59) Heterologous hosts must be genetically manipulatable and generally contain both sufficient fluxes of metabolic precursors as well as an appropriate environment for protein folding and expression. (60)

There are challenges that are specific to megasynthase heterologous expression, particularly in unrelated organisms like *E. coli.* These challenges include metabolic burden inhibiting host cell growth, cytotoxicity of metabolites, a high degree of unfolding, and insufficient precursor supply. (61,62) As PKSs/NRPSs are usually large, complex multi-domain proteins that are held together via flexible linkers, (63) and thus they are particularly prone to expression and solubility challenges and/or premature truncation, which is of particular importance when selecting appropriate transcriptional and translational machinery. These complexities result in an unwieldy number of parameters that can be varied when selecting an appropriate expression and screening strategy. Although a perfect system for heterologous host expression likely does not exist, determining the advantages and disadvantages of different expression systems will enhance the ability to generate engineered secondary metabolites. For these reasons, it is necessary to investigate different protein expression options to better identify how to express megasynthases from the actinomycetal origin and to appraise the pros and cons of different expression choices for different experimental applications.

1.4. Lysate-Based Cell-free Protein Synthesis for Heterologous Pathways Enzymes

In general, it is challenging to profile biosynthesis genes in a high throughput fashion due to their large size and complex structure. The modular organization of PKSs/NRPSs lends itself to engineered biosynthesis where the mixing and matching of domains can generate new molecules. Only a limited set of constructs can be feasibly made under most experimental circumstances, and profiling multiple different coding sequences with different regulatory features is time-consuming (especially if proteins are expressed and purified to homogeneity), which can hinder the rapidity of screening. (64) To overcome these challenges, cell-free protein synthesis (CFPS) or *in vitro* system has emerged as a strategy to prototype expression as well as undergo biomanufacturing strategies that have limitations in cellular contexts for reasons like toxicity. (65)(66) CFPS systems are often divided into two main approaches, which are crude cell lysates/extracts supplemented with appropriate precursors and cofactor recycling systems and the *in vitro* TX-TL PURE system. The crude cell lysates are generated from cell growth and then are supplemented with cofactors and other necessary biological components for *in vitro* transcription and translation (TX-TL). Whereas TX-TL PURE system or PURExpress (67–70) is a system that has taken the minimal recombinant elements required for cell-free expression and is thus less complex than lysate-based systems. The lysate-based system has the advantage over the PURExpress system in that it allows for accessory genes (e.g., ones relevant to posttranslational modification) as well as chaperonins. However, it has the disadvantage that it is more complex and has the potential for more sources of background that can interfere with signal output and/or cofactor recycling. (71,72) The lack of cellular survival objectives bypasses challenges with *in vivo* expression, and thus the protein expression process is shortened significantly (from 1-2 weeks in living cells to 1-2 days) (**Figure 1.2**). (61,73,74)

Recently, there has been rising interest in developing high-throughput profiling of megasynthase genes using *in vitro* platforms such as in cell-free expression (CFE) systems (62,74) that include both lysate-based systems and the PURExpress system. In addition to the advantages in terms of production timeframes, cell-free affords the ability to decouple limitations that exist within cell-based systems including rapidly increasing DNA concentration as well as the introduction of non-natural components like RNA polymerases, and noncanonical amino acids, all without requiring extensive cell engineering. (73,75,76) Thus, this is additionally appealing for megasynthase characterization and engineering because it can accelerate design-build-test-learn (DBTL) cycles. To these ends, megasynthase proteins have recently been expressed in both lysate-based (77,78) (79) (80) and PURExpress (81) cell-free systems. These foundational studies can be built upon to develop more sophisticated pipelines for PKS/NRPS prototyping of valuable small molecules from inexpensive feedstocks like glucose in the future. (82)

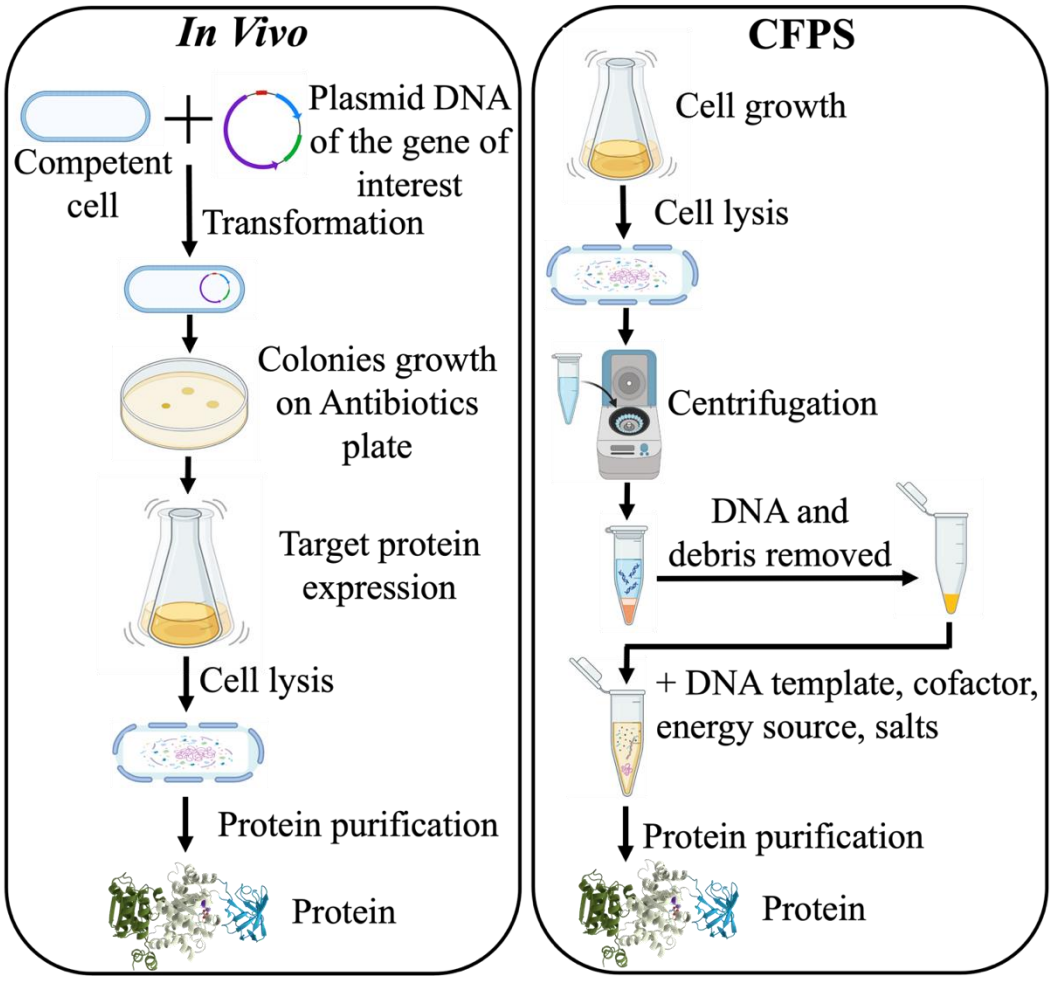


Figure 1.2. Comparison of *in vivo* recombinant DNA protein expression and cell-free protein synthesis (CFPS) methods.

1.5. Scope of Dissertation

It is significant to be able to express complex biosynthetic genes such as NRPSs because if an expression cannot even be achieved, then all other downstream strategies to generate a heterologous host for metabolite production (such as metabolic engineering to optimize precursor flux) are not viable. This dissertation aims to expand the current toolkit to express biosynthetic genes and to better understand the pros and cons of different expression choices for different experimental applications. The focus is on 1) probing the effect of refactoring an NRPS for heterologous expression in *E. coli*, 2) optimizing the lysate CFE TX-TL machinery for the expression of megasynthase proteins, and 3) profiling expression choices for biosynthetic proteins *in vitro* and *in vivo*. The approaches can advance options to express and achieve a high yield of multienzyme heterologous proteins.

1.6. Probing Effects of Refactoring Biosynthetic Megasynthases for Heterologous Expression in *Escherichia Coli*

Even though heterologous expression has been approached efficiently in many different organisms, an ideal "universal host" for protein heterologous expression does not exist. *Escherichia coli* is typically the first choice for heterologous protein expression when performing biochemical studies seeking to generate a fair number of overexpressed proteins because of its fast growth, (83) facile transformations with exogenous DNA, (84) and genetic tractability. *E. coli* has been used as a workhorse organism for molecular biology because it is by far the most understood, characterized, and genetically tractable organism. However, as most of Actinomycetes have ~75% GC content, which is significantly higher than that of *E. coli*, it often results in poor expression of genes from actinomycetes in *E. coli*. The disparity in GC content between source DNA and host organism correlates to several regulatory differences in transcription and translation that can impact the kinetics of peptide bond formation at the ribosome and often result in issues with expression and folding. In particular, rare codon usages from differing codon biases have shown an impact on the expression of heterologous proteins. (85) For these reasons, proteins of actinomycetal origin are often re-coded for *E. coli* due to the differences in codon bias between these two organisms. However, alterations to a codon usage pattern can significantly affect the amount of proteins synthesized, and the yield of correctly folded proteins, and thus often results in premature truncation and/or poorly folded proteins.(64) Moreover, amplification of high GC genes by PCR is often non-trivial, requiring additives, tailored thermocycling protocols, specialized polymerases, and challenging gene synthesis. Thus, if a gene cannot be amplified from genomic DNA, an alternative way to get access to the native coding sequence often is recoding, as most commercial gene synthesis vendors commonly do not generate genes with such large sizes and high GC content. However, decreased transcript stability in the cell has been shown when changing a gene to a synonymous synthetic gene. (86) While these differences have been noted in other systems, such as work by Burgess-Brown and co-workers on eukaryotes genes, (87) to date the exact

impacts of different expressions on complex multidomain proteins from distantly related prokaryotes (such as NRPSs) are unclear. (88)

Hence, a codon-optimized sequence and a native sequence of a *Streptomyces* protein were heterologously expressed in *E. coli* to reveal the effects of refactoring biosynthetic megasynthases for heterologous expression in *E. coli* (Chapter 2). An NRPS from *Streptomyces lavenduale*, a multidomain "megasynthase" gene that comes from a high GC (72.5%) genome, blue pigment synthetase A (BpsA) (>100 kDa), was used as a model NRPS for the head-to-head comparison of codon-optimized sequences versus a native sequence of proteins of streptomycete origin heterologously expressed in *E. coli*. This work showed that any disruption in co-translational folding from codon mismatch that reduces the titer of indigoidine is explainable via the formation of more inclusion bodies as opposed to compromising folding or posttranslational modification in the soluble fraction. This result supports that any refactoring strategies to improve soluble expression in *E. coli* can be applied without concern about the protein that reaches the soluble fraction being differentially folded.

1.7. Improving Lysate-based CFE Strategies for Megasynthase Proteins

While several classes of proteins have been generated using CFPS, CFPS systems are typically optimized and developed via applying tractable and detectable reporters, predominantly GFP and other fluorescent proteins like mScarlett. (76,89,90) Multienzyme proteins pose more demands to a cell-free system than simple fluorescent protein reporters due to their large size, complex multidomain architecture, posttranslational modification, and cofactor requirements (e.g., ATP) and thus the efforts to express megasynthase enzymes and produce complex metabolites in cell-free are somewhat limited. Despite the successful proof of concept that NRPS can be successfully expressed from the Jewett and Gringiner labs, further optimization of the CFE conditions was likely necessary to further unlock the ability to generate NRPS in cell-free. (91,92) CFPS tends to use GFP as the initial protein for optimization; (91) however, GFP differs substantially from biosynthetically relevant proteins such as NRPSs. For instance, the A domain uses ATP to generate an aminoacyl-AMP intermediate, which could potentially compete with CFPS components (e.g., amino acids, phosphoenolpyruvate (PEP), and magnesium ions). Additionally, with a large protein beyond the size of native *E. coli* proteins, premature truncation can be a concern. (77,78,93) Hence, different reaction conditions are examined to probe the limitations of NRPS CFE (79) (Chapter 3). Reporters were utilized, as many of the methods to evaluate full-length protein translation are low throughput (e.g., SDS-PAGE gels or proteomic evaluation via mass spectrometry) (77,78) This work applied a tetracysteine (TC) peptide tag expressed at the C-terminal of the expressed protein to investigate factors that limit full-length NRPS translation, and consequentially, metabolite production in an *E. coli* lysate-based cell-free system (79) via an exogenously added organoarsenic dye that binds to the TC tag to form a fluorescent complex, which was

previously established in the PURExpress cell-free system. (71) This technique can be used to quickly measure the full-length expression of megasynthase under different CFE reaction conditions. The blue pigment synthetase A (BpsA) (>100 kDa) was used as a model NRPS to optimize CFE conditions via this technique. The CFE-optimized conditions for BpsA also showed improvement in the expression of other NRPSs, demonstrating general applicability.

1.8. Profiling Recombinant Protein Expression Strategies for Actinomycetal Origin Proteins in a Lysate-Based, Cell-free System

Recombinant proteins are important tools in microbial systems as they have a wide range of applications in the biomanufacturing industries. However, optimal conditions to express a protein of interest heterologously do not exist. Therefore, it is critical to establish a production system for proteins in a manner that is high-yield and cost-effective. Choices and design of plasmids play an important role in minimizing metabolic burden and increasing yield for recombinant proteins. Hence, to achieve a more effective system, an important parameter to consider is the characteristic of genetic elements (e.g., coding sequence and/or promoter). Besides time-consuming procedures, the bottleneck of conventional cell-based methods of protein expression often is cell toxicity. To bypass this, cell-free expression (CFE) platforms derived from crude cell extracts can be used to prototype plasmid constructs for *in vivo* enzyme production, expediting genetic parts characterization from weeks to days. CFE is a simple open system that allows direct manipulation of transcription, and translation, without the use of living cells; hence, it is easier to control via varying parameters. While proven for robustly synthesized proteins, lysate-based CFE's utility for prototyping expression of more complex proteins, such as biosynthetic pathways remains unclear. Hence, a type III polyketide synthase from *Streptomyces griseus*, RppA, which catalyzes the formation of the red pigment flaviolin, was used as a reporter to investigate biosynthetic gene clusters (BGC) refactoring techniques (Chapter 4). This work synergistically tunes promoter and codon usage to improve flaviolin production from cell-free expressed RppA. The utility of cell-free systems for prototyping these refactoring was then used to assess the tactics prior to the implementation in cells. Refactoring promoters and/or coding sequences via CFE can be a valuable strategy to rapidly screen for catalytically functional production of enzymes from BCGs. Establishing the utility of lysates for prototyping PKS expression will accelerate the construction of valuable polyketide-forming microbial cell factories, diversifying access to natural product scaffolds.

1.9. Organization of Dissertation

Chapter 2 is adapted from the article "Expression of Blue Pigment Synthetase A from *Streptomyces Lavenduale* Reveals Insights on the Effects of Refactoring Biosynthetic

Megasynthases for Heterologous Expression in *Escherichia coli*" originally published in the *Protein Expression and Purification* in 2023. (94) Chapter 3 is adapted from the article "Applying Reporters to Investigate and Optimize Lysate-Based Cell-Free Expression for Nonribosomal Peptide Production" which was published in 2023 in *ACS Synthetic Biology*. (79) Chapter 4 of this manuscript is adapted from the article "Profiling Expression Strategies for a Type III Polyketide Synthase in a Lysate-Based, Cell-free System" which is currently under review for publication as a research article in *Scientific Reports*, and available as a preprint deposited to biorxiv in 2023. (95) The Appendix section at the end of each chapter includes the Supplementary Information associated with each article. The References for each chapter are located at the end of this dissertation.

CHAPTER 2

EXPRESSION OF BLUE PIGMENT SYNTHETASE A FROM STREPTOMYCES LAVENDUALE REVEALS INSIGHTS ON THE EFFECTS OF REFACTORING BIOSYNTHETIC MEGASYNTHASES FOR HETEROLOGOUS EXPRESSION IN ESCHERICHIA COLI

A version of this chapter was originally published by Tien T Sword, J William Barker, Madeline E. Spradley, Yan Chen, Christopher J Petzold, and Constance B Bailey:

Sword, T. T., Barker, J. W., Spradley, M., Chen, Y., Petzold, C. J., and Bailey, C. B. (2023) Expression of blue pigment synthetase a from Streptomyces lavenduale reveals insights on the effects of refactoring biosynthetic megasynthases for heterologous expression in Escherichia coli. Protein Expr. Purif. 210, 106317.

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

2.1. Abstract

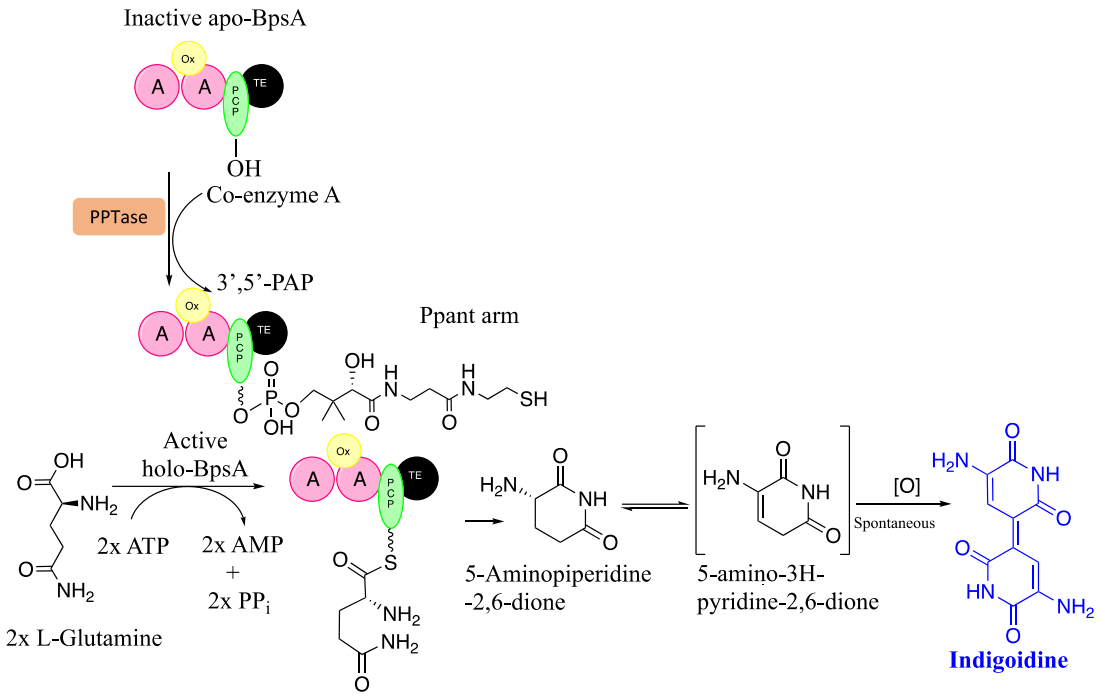
High GC bacteria from the genus *Streptomyces* harbor expansive secondary metabolism. The expression of biosynthetic proteins and the characterization and identification of biological "parts" for synthetic biology purposes from such pathways are of interest. However, the high GC content of proteins from actinomycetes in addition to the large size and multi-domain architecture of many biosynthetic proteins (such as non-ribosomal peptide synthetases; NRPSs, and polyketide synthases; PKSs often called "megasynthases") often presents issues with full-length translation and folding. Here we evaluate a non-ribosomal peptide synthetase (NRPS) from *Streptomyces lavenduale*, a multidomain "megasynthase" gene that comes from a high GC (72.5%) genome. While a preliminary step in revealing differences, to our knowledge this presents the first head-to-head comparison of codon-optimized sequences versus a native sequence of proteins of streptomycete origin heterologously expressed in E. coli. We found that any disruption in co-translational folding from codon mismatch that reduces the titer of indigoidine is explainable via the formation of more inclusion bodies as opposed to compromising folding or posttranslational modification in the soluble fraction. This result supports that one could apply any refactoring strategies that improve soluble expression in E. coli without concern that the protein that reaches the soluble fraction is differentially folded.

2.2. Introduction

Bacteria from the order Actinomycetales, especially those from the genus *Streptomyces* are some of the most prolific producers of bioactive natural products. Indeed, over half of the clinically used antibiotics arise from *Streptomyces* and related Actinomycetes (96,97). Many of these natural products come from classes such as

polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs) which generate the polyketide and non-ribosomal peptide metabolites, respectively. PKSs and NRPSs are megadaltons in size (often termed "megasynthases") (98–100). The large size of PKSs and NRPSs combined with their high GC genomes (~75% GC) often results in poorly understood issues with protein expression and folding in heterologous hosts, especially those that are evolutionarily distinct and more moderate in GC content (101). While numerous investigations have focused on the improvement of metabolic flux (102–105), the availability of phosphopantetheinyltransferases (PPtases) to provide sufficient posttranslational modification (106), and genome minimization (107–109) to remove potential drains on metabolic resources, such as host optimization efforts will be fundamentally limited if the key bottleneck remains to create full length, well folded and catalytically active protein. *Streptomycetes* harbor biosynthetic genes involved in secondary metabolism in general, and particularly PKSs and NRPSs have had extensive investigation with the goal of creating engineered products including medicinal agents (102,104) as well as commodity and specialty chemicals (11,110,111). None of these synthetic biology applications are feasible without the critical step of biochemical characterization of biosynthetic proteins.

Well-established heterologous expression systems exist for various *Streptomyces* sp. (112–115), however, they typically remain a tool of last resort for the application of production of adequate protein for overexpression and biochemical characterization, which is a fundamental step for characterizing synthetic biology parts. This is due to their slow growth curves, less than ideal growth properties (e.g. mycelial clumping), and typical requirement of transfer of genetic material through specialized techniques (e.g. intergenic conjugation or lengthy protoplast preparations) results in less than ideal tractability for isolation and purification of protein (112). Despite the vast evolutionary differences between *E. coli* and *Streptomyces* and related actinomycetes, the unparalleled genetic tractability, as well as the rapid doubling time and facile growth conditions, still fuels motivation for expressing high GC and large proteins within *E. coli*. While *Streptomyces* species might be all-around better hosts for metabolite production in many settings (116), the downsides of expression for protein overexpression still typically weigh the scale towards optimizing expression in *E. coli* for this key application. With *E. coli*, due to the large differences in GC content between *Streptomyces* sp. and *E. coli*, this means that it is not always obvious whether an expression is improved by maintaining the native coding sequence or by performing codon optimization to better match the codon usage of *E. coli* when considering the need for highly expressed appropriately folded soluble protein. Additionally, the native coding sequence may be intractable to generate as a synthetic gene due to gene synthesis companies' lack of capacity to produce high GC DNA sequences, necessitating a synthetic gene that uses synonymous codons that are more synthetic accessible if the organism is not readily cultured and/or the gene is challenging to clone from genomic DNA.



Scheme 2.1. Production of indigoidine by the single-module NRPS BpsA.

Schematic diagram showing apo-bpsa activated to holo-bpsa via attachment of a phosphopantetheinyl prosthetic group derived from Co-enzyme A, mediated by a phospopantetheinyltransferase. Two molecules of L-glutamine are converted by holo-bpsa into the easily detectable blue pigment indigoidine. Bpsa consists of an adenylation (A) domain with an oxidase (Ox) domain present between subdomains of the A-domain, a peptidyl carrier protein (PCP) domain, and a thioesterase (TE) domain.

Despite the mechanistic knowledge that codon usage has an evolutionary basis to regulate the rate of translation and co-translational folding (117–120), there are contradictory anecdotal reports of success and failure of codon-optimized constructs among members of the natural products enzymology and synthetic biology communities. Even though the genetic code is universal, the usage frequencies of synonymous codons can differ substantially between organisms. Codon bias is when there is a preferred usage of one codon over another for the same amino acid. When codon usage differs in a heterologous host, codons that are rarely used in *E. coli* such as CUA (Leucine); AGA, AGG (Arginine); AUA (Isoleucine); CCC, (Proline); GGA, GGG (Glycine) essentially regulates the expression of different endogenous protein (AGG, CCC, GGA, and GGG are common in *Streptomyces sp.)*.(87,121) The level of reduction of protein expression in *E. coli* corresponds with the relative positions of these rare codons in genes. Due to the lack of cognate tRNAs corresponding with rare codons in the host organism, the expression of heterologous proteins is limited. With such extreme GC bias, deciding whether to codon optimize or not is not always clear to the researcher. While some head-to-head comparisons of *E. coli* codon-optimized constructs compared to natively coded constructs exist for eukaryotic proteins (87), few analogous experiments have not been performed for high GC prokaryotes. To date, the only comparison of codon-optimized *Streptomyces* constructs to a natively coded construct has been for a ß-glucanase enzyme which indicated small but measurable differences in protein folding that affected protein activity between synonymous coding sequences(122). Because there is high interest in expressing biosynthetic genes for synthetic biology purposes from such metabolically gifted bacteria, we sought to do a head-to-head comparison of a model megasynthase with the supplementation of tRNA for rare codon in *E. coli*. For our model, we chose blue pigment synthetase A (BpsA) from *S. lavendulae* which produces the blue pigment, indigoidine (123). BpsA is an example of type 1 non-ribosomal peptide synthetase (NRPS) (124), which is a large, multi-domain protein with flexible linkers and an example of a protein that provides unique challenges to heterologous expression among relevant biosynthetic proteins from *Streptomyces* that we sought to interrogate (**Scheme 2.1)**.

2.3. Methods

2.3.1. Strains and plasmids

Chemically competent *E. coli* DH5α cells, *E. Coli* BL21(DE3) Rosetta were purchased from Novagen, USA. BL21(DE3) Rosetta strain contains a plasmid harboring tRNA genes for the following rare codons: AGG, AGA, AUA, CUA, CCC, and GGA on a chloramphenicol-resistant plasmid. Chemically competent *E. coli* BAP1 was obtained from Prof. Christopher Boddy **(**University of Ottawa, Ottawa, Canada). *E. coli* BAP1 essentially is a derivative of the commonly used of *E. coli* BL21(DE3) strain, which has the promiscuous PPTase *sfp* to provide the required phosphopantetheinyl post-translational

modification.(106) Hence, with present of *sfp* allows BAP1 to generate phosphopantetheinylated active *holo*-ACPs and PCPs. Native *bpsA* in the pET28 vector was obtained from Prof. Michael Burkart (University of California San Diego) (125). The *E. coli* codon-optimized sequences of *bpsA* were designed using Integrated DNA Technology's (IDT, USA) codon optimization tool and ordered as two gene blocks (part 1 and part 2) from Integrated DNA Technology's (USA) (supplementary information). The construct was cloned using GoldenGate cloning kit (NEB, USA) and included an N-terminal His_6-tagged and a stop codon (TAA). Primers were designed with the J5 algorithm (126). Primers used in this study are listed in **Table S2.2**. GoldenGate cloning kit, DNA ladder (1kb), and protein ladder were purchased from New England BioLabs, USA. PrimeSTAR GXL Premix was purchased from Takara Bio Inc., USA. PCR cleanup kits, miniprep kits, and gel recovery kits were purchased from Zymo Research, USA. Primers were obtained from Thermofisher, USA. Chemicals required for SDS-PAGE, purification, Luria Bertani (LB) broth, and agar were purchased from Thermofisher. Pre-made protein gel was purchased from NuSep, USA. All constructs were confirmed by Sanger sequencing.

2.3.2. Expression and Purification of BpsA

Expression of BpsA was conducted using either *E. coli* BAP1(106) as a host strain or *E. coli* BAP1 transformed with the tRNA complementation plasmid isolated from *E. coli* BL21(DE3) Rosetta (Novagen, USA). Cultures were grown in LB broth (Miller) supplemented with appropriate antibiotics (kanamycin at 50 µg/mL and 25 µg/mL chloramphenicol as appropriate). Overnight seed cultures were grown in 25 mL LB broth and kanamycin (50 µg/mL) inoculated with a single colony at 37 °C, shaking at 210 rev/min. 2L of expression cultures were inoculated from these cultures in a ratio of 1:100 and incubated at 37 °C shaking at 210 rev/min until an OD_{800} of 0.3-0.4 was reached. Note that OD_{800} rather than the more typical OD_{600} was used to monitor growth to ensure that interference from the signal from indigoidine would not interfere with measurements of cellular turbidity. This strategy was adapted from Beer and coworkers (127) and a reproduction of the correlation between the OD_{600} and OD_{800} is shown in (**Figure 2.1**). The temperature was then lowered to 16 °C and cooled for 15 minutes prior to induction by the addition of IPTG to a final concentration of 0.5 mM. The cultures were then incubated for ~20 hours at 16 °C prior to harvesting by centrifugation (4000 x g, 45 min at 4 °C).

Cell pellets were resuspended in a wash buffer (5 mM imidazole, 0.5 M NaCl, 10% v/v glycerol, 50 mM sodium phosphate, pH 7.8) and lysed by sonication (3x 30 min on, 1 min off). After sonication, the lysate was clarified via centrifugation (11,000 x g, 45 min, 4 °C) and purified via IMAC on an AktaPure system (Cytiva, USA) using a 5 mL HisTrap column (Cytiva, USA). Protein was eluted with elution buffer (400 mM imidazole, 0.5 M NaCl, 10% v/v glycerol, and 50 mM sodium phosphate, pH 7.8) with the gradient from 20-60%, 1 column value (CV), 60-100, 1 CV, and hold at 100% for 3 CV. After purification,

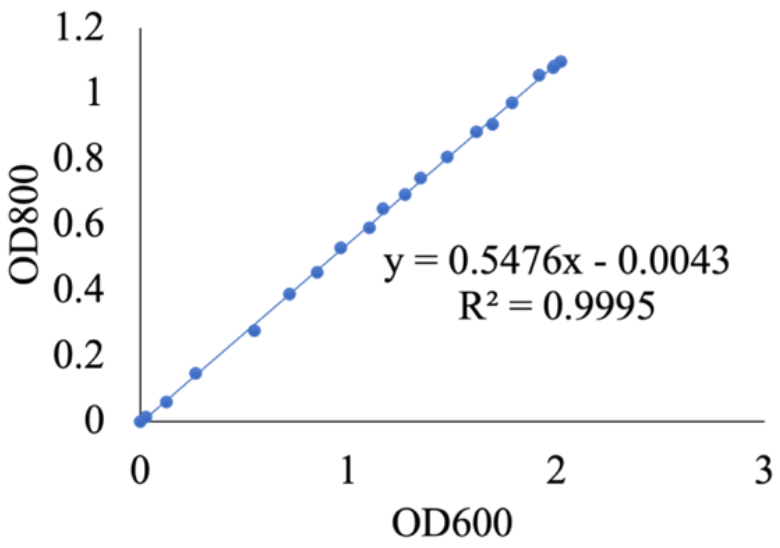


Figure 2.1. Correlation of the optical density of BAP1 between OD800 and OD600 measurements.
The accuracy of the R2 shows that we can measure the growth in this study at OD800.

the collected fractions were dialyzed (10K MWCO dialysis tubing Thermofisher) for 2 passes for a minimum of six hours into storage buffer (50 mM sodium phosphate pH 7.8, 100 mM NaCl, 10% v/v glycerol). The protein was then concentrated using a 100 K MWCO concentrator (Thermofisher, USA). Purity was assessed via SDS-PAGE using 4-20% pre-made gels (NuSep, USA).

2.3.3. Indigioidine Purification and Preparation of the Indigoidine Standard Curve

A 10 mL seed culture containing BAP1 harboring *E. coli* coded *bpsA* was grown from a fresh single colony overnight (37 °C, 2010 rev/min) in LB medium supplemented with 50 µg/mL kanamycin. After ~20 hours, 1L of culture was inoculated with the overnight seed culture in a ratio of 1:100 and incubated at 37 °C shaking at 210 rev/min. At an OD_{800} of 0.3-0.4, the overnight culture was induced with 0.5 mM IPTG, grown at 16 °C for 20 h, and then lyophilized for 36 hours. The dry cell mass was then washed with three rounds of water, methanol, ethanol, isopropanol, and hexanes to remove metabolites, salts, and proteins. Finally, the product was dried for ~one week under vacuum. Purity was verified via ^{1}H NMR: (400 mHz, D_6MSO) δ 11.30 ppm (s, NH), 8.18ppm (s, CH), and 6.46ppm (s, NH_2) which is in agreement with the previous literature report (128,129). Afterward, 0.5 mg of dry indigoidine was dissolved in 1 mL DMSO. This solution was then serially diluted to six different concentrations (0.01, 0.025, 0.5, 0.1, 0.2, and 0.25 mg/mL). 200 µL of the solution was added to a 96-well plate in triplicate to measure A_{595} to generate a standard curve.

2.3.4. Measurement of Titer

E. coli BAP1 transformed with the appropriate plasmid(s) were grown in 250 mL Erlenmeyer flasks containing 40 mL of LB broth (~15% filling volume) supplemented with 50µg/ml kanamycin (37˚C, 210 rpm). The experiment was performed in triplicate. When the OD_{800} value reached 0.3-0.4, 500 µM IPTG was added to induce the expression of BpsA. After the induction, the fermentation broths were incubated at 16˚C and 210 rpm for 45 min, and then L-glutamine was added with the final concentration at 1.5 g/l. The broths were maintained growing for an additional 24 hours. The cultures were then harvested to measure the titers of indigoidine. 1 mL of fermentation broth was centrifuged (11000 x g for 30 minutes), the supernatant was discarded, and the cell pellets were washed with 1mL of water, ethanol, methanol, isopropanol and then dissolved in 1ml of DMSO by pipette. The insoluble component was removed by centrifugation (8000 xg, 20 minutes). The absorption value of the DMSO solution was measured at 595 nm triplicated. The titer of indigoidine was then calculated based on the standard curve of pure indigoidine.

2.3.5. Thermal shift assay

A Thermal shift assay was used to determine the melting temperature (T_m) of BpsA using SYPRO orange (130). Each BpsA expressed under each expression condition was purified to homogeneity. Triplicate reactions of three separate preparations were established for the assay. In a 96-well plate, reaction conditions were as follows 20µL storage buffer (50 mM phosphate pH7.8, 100mM NaCl, 10%(v/v) glycerol), 5 µL 50x SYPRO orange dye, 15 µL Millipore water, and 10 µL of protein with the concentration between 0.5-7 mg/mL. Lysosome (1mg/ml) was used as the positive control. Briefly, each well was measured from 20 °C to 98 °C with the rate was 1 °C. The corresponding of the lowest measurement of the derivative data used to determine T_m of BpsA (supplementary information).

2.3.6. Proteomics Analysis

50 µg of each purified protein sample was precipitated using the established acetone precipitation method (131). Proteins were resuspended in 100 mM ammonium bicarbonate buffer supplemented with 20% methanol, followed by reduction using 5 mM tris 2-(carboxyethyl) phosphine (TCEP) for 30 min at room temperature, and alkylation with 10 mM iodoacetamide (IAM; final concentration) for 30 min at room temperature in the dark. Overnight digestion with trypsin was accomplished with a 1:50 trypsin: total protein ratio. The resulting peptide samples were analyzed on an Agilent 1290 UHPLC system coupled to a Thermo scientific Obitrap Exploris 480 mass spectrometer for the discovery of proteomics (132). Briefly, 20 µg of tryptic peptides were loaded onto an Ascentis® (Sigma–Aldrich) ES-C18 column (2.1 mm × 100 mm, 2.7 µm particle size, operating at 60°C) and were eluted from the column by using a 10-minute gradient from 98% buffer A (0.1 % FA in H_2O) and 2% buffer B (0.1% FA in acetonitrile) to 65% buffer A and 35% buffer B. The eluting peptides were introduced to the mass spectrometer operating in positive-ion mode. Full MS survey scans were acquired in the range of 300-1200 m/z at 60,000 resolutions. The automatic gain control (AGC) target was set at 3e6, and the maximum injection time was set to 60 ms. The top 10 multiply charged precursor ions (2-5) were isolated for higher-energy collisional dissociation (HCD) MS/MS using a 1.6 m/z isolation window and were accumulated until they either reached an AGC target value of 1e5 or a maximum injection time of 50 ms. MS/MS data were generated with a normalized collision energy (NCE) of 30, at a resolution of 15,000. Upon fragmentation precursor ions were dynamically excluded for 10 s after the first fragmentation event. The acquired LCMS raw data were converted to mgf files and searched against the latest UniProt *E. coli* protein database supplemented with BpsA and other common contaminant protein fasta sequences using Mascot search engine version 2.3.02 (Matrix Science). Phosphopantetheine modification to L-serine was defined as a variable modification in addition to other common structural modification parameters, such as carbamidomethyl and oxidized methionine. The resulting search results were filtered and analyzed by Scaffold v 5.0

(Proteome Software Inc.). The quantitative report of phosphopantetheine-modified peptide was analyzed by Skyline v 22.2 (University of Washington).

2.4. Results and Discussion

2.4.1. Expression and purification of BpsA

To test the hypothesis that the difference in coding sequence might impact BpsA stability via disruptions to co-translational folding, we cloned two synonymously coded constructs of *bpsA* into the commonly used expression vector, pET28a. Because of the T7 promoter, high copy number origin, and lac-inducible operon, this is typically a commonly used vector for overexpression and purification for biochemical characterization and is usually the first choice for the application of biochemical characterization. First, we used a coding sequence that was cloned directly from genomic DNA(125) of *S. lavendulae* as the native coding sequence. Next, we ordered *bpsA* as a gene block from Integrated DNA Technologies using their Codon Optimization Tool to create a synonymously coded *E. coli* codon-optimized construct of *bpsA*. The natively coded *bpsA* gene had a GC content of 68% whereas the *E. coli* codon-optimized gene had a GC content of 51%. The overall genome of *Streptomyces lavenduale* has a GC content of 72.5%, thus the GC content of *bpsA* is lower than a typical gene from *Streptomyces lavenduale*.

We measured purified protein as a proxy for how much fully translated, soluble protein was present. The electrophoretic analysis of the complete purification process as well as the purity was assessed via SDS-PAGE (**Figure 2.2, Appendix Figure S2.1**). The expression of both plasmids was performed in *E. coli* BAP1, a derivative of the commonly used BL21(DE3) strain specifically designed for PKS and NRPS expression a copy of the promiscuous phosphopantetheinyl transferase, *sfp* is integrated into its genome (PKS and NRPS expression require phosphopantetheinylation for catalytic activity). We saw a lower purified yield of protein with the native codon construct than with the *E. coli* codon-optimized construct (1.4±0.1 mg/L versus 2.1±0.17 mg/L respectively, **Table 2.1**).

2.4.2. tRNA Supplementation for Rare Codons in *E. coli*.

Because the *E. coli* codon-optimized construct showed a higher yield of purified protein than the native codon construct, we decided to pursue another common strategy to improve the expression of genes with a codon usage that differs from *E. coli*, using tRNA complementation plasmids. The commercially available BL21(DE3) Rosetta strain contains a plasmid harboring tRNA genes for the following rare codons: AGG, AGA, AUA, CUA, CCC, and GGA on a chloramphenicol-resistant plasmid. To ensure that we retained phosphopantetheinylation, we isolated this plasmid from the commercially

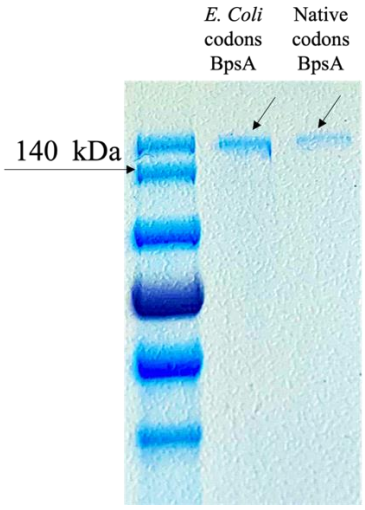


Figure 2.2. Two constructs of BpsA were purified to homogeneity with a molecular mass of 145 kDa.
The expression of *E. coli* codon-optimized bpsA and native-coded bpsA was performed in *E. coli* BAP1, which is a subsequence strain of the commonly used BL21(DE3). Because PKS and NRPS expression requires phosphopantetheinylation for catalytic activity, sfp is integrated into BL21(DE3) genome to create BAP1 strain for the expression of PKSs and NRPSs. This figure indicates the purity of each construct purified to homogeneity.

available BL21(DE3) Rosetta strain and transformed it into *E. coli* BAP1. When supplemented with the rare tRNA complementation plasmid, the yield of purified protein becomes comparable within prep-to-prep variability, with the native coding sequence slightly higher yielding (5.2±0.38 mg/L versus 4.8±0.48 mg/L for native coding and E. *coli* codon-optimized constructs, respectively) (**Table 2.1)**. However, the levels of insoluble protein become more comparable between the two conditions when qualitatively comparing via SDS page gel evaluating the soluble versus insoluble fractions (**Figure 2.3**). The additional tRNA not only improves the yield of the native construct but also improves the yield of the *E. coli*-coded construct. This can be explained that the Rosetta plasmid provides a rich source of tRNA in the translation process, hence the yields are increased for both constructs. Interestingly, even though there is an increase in the insoluble protein of the E. *coli* codon-optimized constructs when adding the Rosetta plasmid, the yield of purified protein also increased. This means that the total proteins that are expressed are significantly increased along with the inclusion bodies, and the increase in recovered soluble protein is significant (see **Figure 2.4**).

2.4.3. Measurement of Indigoidine Titer

As an output for the functional activity of BpsA in a context relevant to metabolite production, we measured the titer of indigoidine to better understand the functional consequences of codon optimization. A standard curve for purified indigoidine was established (**Appendix Figure S2.3**). We saw a distinct difference in titer with approximately 2-fold more indigoidine production in the *E. coli* codon-optimized construct than in the native construct (0.41 ± 0.03 mg/mL and 0.21± 0.06 mg/mL, respectively for the *E. coli* codon-optimized construct versus the native construct). (**Figure 2.6, Table 2.2**). This result also correlates with purified protein yield (**Table 2.2**). These trends are largely explainable by the fact that more protein was observed in the insoluble fraction as inclusion bodies via qualitative examination of the soluble versus insoluble fractions via SDS-PAGE (**Figure 2.3**). When supplemented with the rare tRNA complementation plasmid, the titers became comparable between the two coding constructs (0.49 mg/mL) (**Figure 2.6)**. Because the levels of insoluble protein become more comparable between the two conditions when supplemented with Rosetta plasmid when qualitatively comparing via SDS-PAGE (**Figure 2.3**). When supplemented with the rare tRNA complementation plasmid, the titers became comparable between the two coding constructs (0.49 mg/mL) (**Figure 2.6)**. Because the levels of insoluble protein become more comparable between the two conditions when supplemented with Rosetta plasmid when qualitatively comparing via SDS page gel evaluating the soluble versus insoluble fractions (**Figure 2.3**), which explains why both the titer of iindigoidine, and yield of purified protein were comparable for each coding sequence.

Table 2.1. Tm values, purified protein yields, and indigoidine titers for each expression condition of BpsA

Construct	Tm (˚C)	Purified Protein Yield (mg/L)	Titer of indigoidine (mg/ml)
E. coli codon-optimized bpsA	41.3 ± 0.06	2.1 ± 0.17	0.41 ± 0.03
Natively coded bpsA	41.3 ± 0.12	1.4 ± 0.10	0.21 ± 0.06
E. coli codon optimized bpsA expressed with the Rosetta plasmid	41.2 ± 0.10	4.8 ± 0.48	0.49 ± 0.005
Natively coded bpsA expressed with the Rosetta plasmid	41.1 ± 0.05	5.2 ± 0.38	0.49 ± 0.003

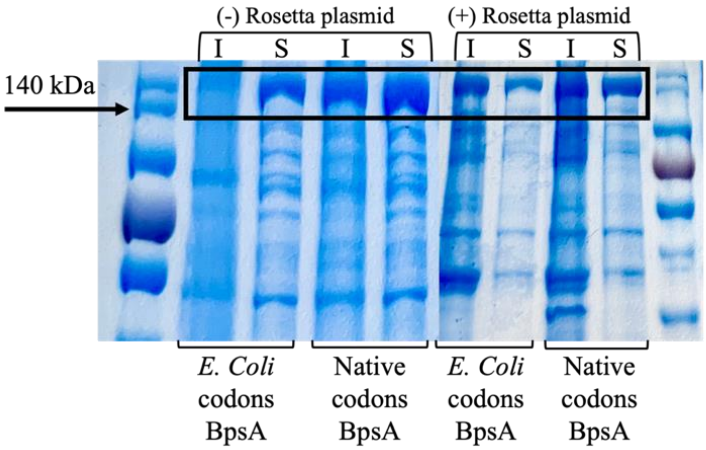


Figure 2.3. Solubility of BpsA.
BpsA expressed in E. coli BAP1 with and without the addition of tRNA from BL21(DE3) Rosetta. Without the Rosetta plasmid, the native coding sequence is less soluble than the E. coli codon-optimized construct. With the Rosetta plasmid, the expression between the two constructs is improved.

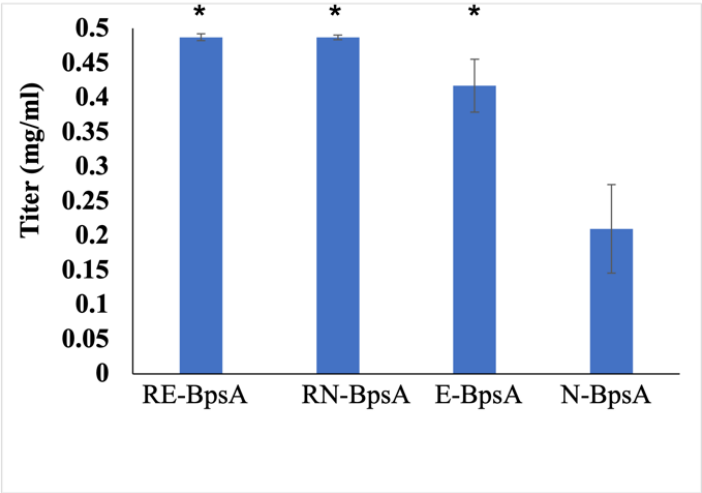


Figure 2.4. Titer of indigoidine.
Without the Rosetta plasmid, the native coding sequence produced less indigoidine than the E. coli codon-optimized construct. With the Rosetta plasmid, the production of indigoidine becomes comparable between the two constructs. Evaluated with a Student's t test. *p-value <0.01.

2.4.4. Thermal Melting Shift Assay

To determine if the soluble, fully translated protein that we were able to purify to homogeneity (**Figure 2.2**) was identically folded, we measured the thermal melting point using the SYPRO orange thermal shift assay (133). All four expression conditions led to a thermal melting point of 41° C (**Table 2.2**). Because the thermal melting points were identical between the four expression conditions, this suggests that any differences between tRNA pool match and issues with co-translational folding lead to the formation of inclusion bodies, but the protein that is well folded enough to remain in the soluble fraction is folded appropriately. Consequently, we can conclude that provided that the protein reaches the soluble layer, it is equally well folded regardless of strategy used to promote increased soluble expression. Curiously, this result is different than what was found by Pradeep and coworkers' experiments investigating synonymously coded Beta glucanase enzymes from *Streptomyces althiiotiicus* TBG-MR17 and *Streptomyces cinereoruber* subsp. *Cinereoruber* TBG-AL3 appeared to have minor but measurable differences in thermal stability between natively coded constructs(122). This difference in results between different classes of proteins from related organisms indicates that more work needs to be done to understand the effects of recoding genes from high GC prokaryotes. It is possible that the large size of BpsA may explain some of these differences, but further investigation needs to be done before we can determine this conclusively.

2.4.5. Posttranslational modification of BpsA

Phosphopantetheinylation is a post-translational modification process, which is important for activation of NRPSs as the peptidyl carrier protein inactive without a posttranslational modification to a conserved serine residue. (134) Because we observed indigoidine production and similar T_m values among all expression conditions, we did not suspect to observe differences in the degree of phosphpantetheinyl posttranslational modification. However, to confirm, a phosphopantetheinyl ejection assay was performed to determine the ratio of *holo/apo*-protein on the PCP domains. This assay was originally developed by Dorrestein and coworkers and was optimized by the Keasling lab. (135,136) Four proteins were purified via the His-tag purification method described above and 20 µL of the purified sample was used to measure their phosphopantetheinylation ratio. Targeted tandem mass spectrometry methods for the phosphopantetheine ejection assay were generated by the Skyline method. (137) The proteomic data were normalized to "global standards", which is a peptide fragment from BpsA (VELDEISLAIENHDWVR) to minimize differences in the BpsA purity between samples. This fragment is identified in all of the samples at high intensity by using the Skyline 'Global Standards' normalization (**Appendix Figure S2.3**). The tryptic PCP fragment containing NSL active site motif is the "parent" ion (ENASVQDDFFESGGN**S**LIAVGLVR, serine that is post-translationally modified with phosphopantetheine is bolded). This fragment carries the

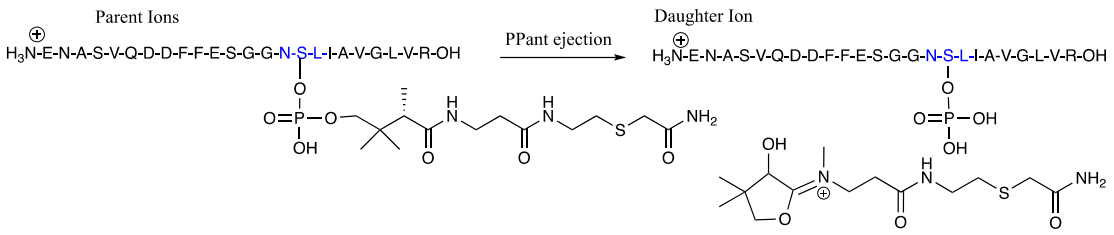


Figure 2.5. PPant ejection fragment after treatment with iodoacetamide during proteomic analysis.
The "parent" ion is the tryptic PCP fragment containing NSL active site motif (S in NSL is the active serine). The phosphopantetheine fragment and acyl group is the 'daughter' ion.

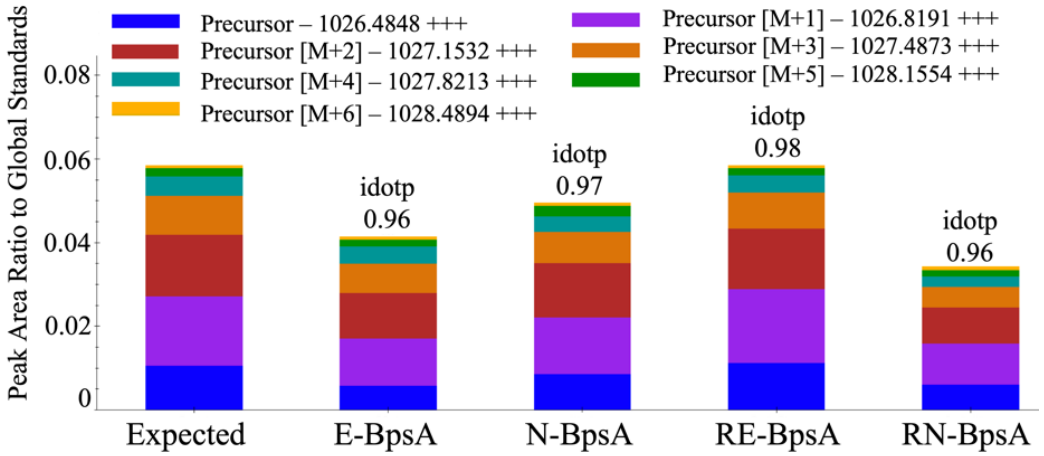


Figure 2.6. Holo-BpsA across all expression conditions:
E-BpsA is *E. coli* codon-optimized BpsA. N-BpsA is natively coded BpsA. RE-BpsA is *E. coli* codon-optimized BpsA with tRNA from *E. Coli* BL21(DE3) Rosetta complementation. RN-BpsA is natively coded BpsA with tRNA from E. Coli BL21(DE3) Rosetta complementation. The "Expected" bar plot represents the predicted isotope distribution of the Holo-peptide as calculated by the software Skyline and corresponds to an idotp value of 1.0.(137) The idotp values of the Holo-peptide for the sample data indicate how close the observed isotope distribution (measured by peak intensity of isotope ions) is to the predicted (Expected) distribution.

phosphopantetheine attachment sites including the acyl-phosphopantetheine. The phosphopantetheine fragment and acyl group is the 'daughter' ion (**Figure 2.5)**.

We found purified BpsA protein across all expression conditions to be present in detectable amounts in the holo-form only without any evidence of detectable apo form, indicating that differences in the percentage of protein that was posttranslationally modified do not explain differences in the titer of indigoidine (**Figure 2.6**). Different native and *E. coli* proteins in the LC-MS analysis across the samples were observed, suggesting that the purity of BpsA is lower in the tRNA complementation group. Since the analysis of the proteomic data focuses on the different states of the BpsA protein and it was detected with high sensitivity in all samples, the absolute abundance differences between the samples do not impact the conclusions drawn from the data. The "Expected" bar plot represents the predicted isotope distribution of the Holo-peptide as calculated by the software Skyline and corresponds to an idotp value of 1.0. The idotp values of the Holo-peptide for the sample data indicate how close the observed isotope distribution (measured by a peak intensity of isotope ions) is to the predicted (Expected) distribution. An idotp value greater than 0.9 is commonly accepted as a cutoff to explain the idotp value. Thus, the idotp values of the Holo-peptide for the sample data indicate how close the observed isotope distribution (measured by peak intensity of isotope ions) is to the predicted (Expected) distribution. While we did see variations in the total peak area of the holo-peptides among different purified sample sources, we believe this is explainable by 1) minor variation in the protein quantification assay used to measure the protein concentration of each sample, and 2) variation of the purity of BpsA enriched via the His-tag purification method. We observed different native *E. coli* proteins in the LC-MS analysis across the samples, suggesting that the purity of BpsA may be lower in the tRNA complementation group. Since the analysis of the proteomic data focuses on different states of the BpsA protein and it was detected with high sensitivity in all samples, the absolute abundance differences between samples do not impact conclusions drawn from the data.

2.5. Conclusions

While there has been much discussion within the literature about how codon usage promotes correct protein folding via evolutionarily tuned co-translational folding events, (118,119,138,139) the details of what the consequences of this are at the protein level are not entirely clear. While one might expect there would be no minor differences in a well-folded protein such as model robust protein such as GFP, (140) megasynthases like BpsA have multiple domains held with flexible linkers with multiple opportunities for partial misfolding that may compromise stability and, consequentially, activity. Because many biosynthetic proteins of interest have more complexity than GFP, BpsA served as an

appropriate proxy for proteins that are of more interest to investigating and engineering biosynthetic pathways. (141)

We found that all expression strategies generated protein with identical Tms, suggesting no detectable difference in stability in the protein that was folded enough to be recovered from the soluble fraction. The differences in protein yield and metabolite titer could be entirely explained by differences in the solubility of translated BpsA, and not protein stability or phophopantetheinylation. As there is evidence that codon usage affects secondary structure formation in E. coli, we wanted to decouple the effects on the solubility expressed protein versus the insoluble protein. (142) The identical stability of BpsA purified from different expression strategies has some important ramifications on refactoring in the context of heterologously expressed megasynthases. This means that regardless of the strategy used to improve expression (synonymous coding or complementation of rare tRNAs), provided that *Streptomyces* NRPS proteins can reach the soluble fraction, it appears to be structurally uncompromised. For proteins that are less tractable than BpsA in the aspect of solubility, which means proteins that are still present as inclusion bodies after codon optimization, suggests a more sophisticated approach such as codon harmonization wherein there is an effort to replace rare codons in the host organism with positions of rare codons in the native organism as opposed to just statistically using the most efficient codon for the new heterologous host as occurs in codon optimization. (143,144) Codon harmonization, a strategy that matches rare codons to align stalling kinetics between the native and heterologous ribosomes may not affect protein folding in a detectable way as long as it is stable enough to reach the soluble fraction. This has implications for exploring better expression strategies for complex biosynthetic proteins more generally so we can have improved biochemical characterization of parts for synthetic biology from Streptomyces and related Actinomycetes.

Appendix

Table S2.1. Plasmid and gene sequences were used in this study.

Name	Sequence
The native coding sequence of BpsA NCBI accession number: BAW81991.1:	atgactcttcaggagaccagcgtgctcgagcccaccctgcaggggaccaccacgttgcccggcctgctcg cccagcgggtggccgaacaccccgaggcgatcgcggtcgcctaccgggacgacaagctcaccttccgc gagctcgcgtccagaagcgcggccctcgccgactacctggagcacctcggtgtctccgccgacgactgc gtcggcctgttcgtcgagccgtcgatcgacctgatggtcggcgcctggggcatcctcaacgccggcgccg cgtacctgccgctgtccccggagtaccccgaggaccggctgcgctacatgatcgagaacagcgagacga agatcatcctggcgcagcagcgcctggtgtcccgtctgcgcgagctcgcgccgaaggacgtcaccatcgt gaccctgcgcgagtccgaggccttcgtccgccccgagggcaccgaggccccggccgcccgcagcgcc cgcccggacaccctcgcgtacgtcatctacacctccggcagcacgggcaagccgaagggtgtgatgatcg agcaccgcagcatcgtcaaccagctcggctggctgcgcgagacctacgcgatcgaccgcagcaaggtca tcctccagaagaccccgatgagcttcgacgccgcccagtgggagatcctctccccggccaacggcgccac cgtcgtcatgggcgccccgggcgtctacgccgaccccgagggcctcatcgagaccatcgtcaagcacaa cgtgaccaccctccagtgcgtcccgacgctgctccagggtctgatcgacaccgagaagttccccgagtgc gtctccctccagcagatcttcagcggtggcgaggccctctcccgcctgctggcgatccagaccacgcagg agatgcccggccgggcgctcatcaacgtctacgggccgaccgagacgacgatcaactcgtcctcgttccc cgtcgaccccgccgacctggacgagggaccgcagtccatctccatcggctccccggtgcacggcaccac gtaccacatccttgacaaggagaccctcaagccggtcggcgtcggtgagatcggcgagctgtacatcggc ggcatccagctggcccgcggctacctgcaccgcgacgacctgaccgccgagcgcttcctggagatcgag ctcgaggagggcgccgagcccgtccgcctgtacaagacgggcgacctcggccagtggaacaacgacgg caccgtgcagttcgccggccgcgccgacaaccaggtcaagctgcgcggctaccgcgtcgagctcgacga gatatccctggcgatcgagaaccacgactgggtccgcaacgccgccgtcatcgtcaagaacgacggccg caccggcttccagaacctgatcgcctgcatcgagctgagcgagaaggaagccgccctgatggaccaggg caaccacggctcccaccacgcgtcgaagaagagcaagctccaggtcaaggcgcagctgtccaacccgg gcctgcgcgacgacgccgagctggccgcccgcccggccttcgacctggagggcgccgagcccacccc cgagcagcgcgcccgggtcttcgcccgcaagacgtaccgcttctacgagggcggcgccgtcacccaggc cgacctgctgggcctgctgggcgccacggtcaccgccggctactcgcgcaaggcggccgacctggccc ccgccgaactcggccagatcctgcgctggttcggccagtacatcagcgaggagcggctcctgccgaagta cggctacgcctccccgggcgcgctgtacgcgacgcagatgtacttcgagctggagggcgtcggcggtctg aagccgggctactactactaccagccggtccgccaccagctcgtcctcatcagcgagcgcgaggccaccg gcaaggccacggcgcagatccacttcatcggcaagaagagcggcatcgagccggtctacaagaacaaca tcctcgaggtcctggagatcgagaccggccacatggtcggcctcttcgagcagatcctgccggcctacggc ctcgacatccacgaccgcgcctacgagccggccgtcaaggacctgctcgacgtcgccgacgaggactac tacctgggcaccttcgagctggtcccgcacgcgggcgcgcgcgacgaccaggccgaggtctacgtccag acgcacggcggaaaggtcgccggcctgcccgagggccagtaccgctacgagaacggcgagctgaccc gcttctcggacgacatcgtcctcaagaagcacgtcatcgcgatcaaccagtcggtgtaccaggccgccagc ttcggcatcagcgtctacagccgcgccgaggaggagtggctgaagtacatcaccctcggcaagaagctcc agcacctgatgatgaacgggctgaacctgggcttcatgtcctcgggctacagctccaagacgggcaaccc gctgccggcctcgcgccgcatggacgccgtcctcggcgccaacggcgtcgacagcgccccgatgtactt

Table S2.1. Continued

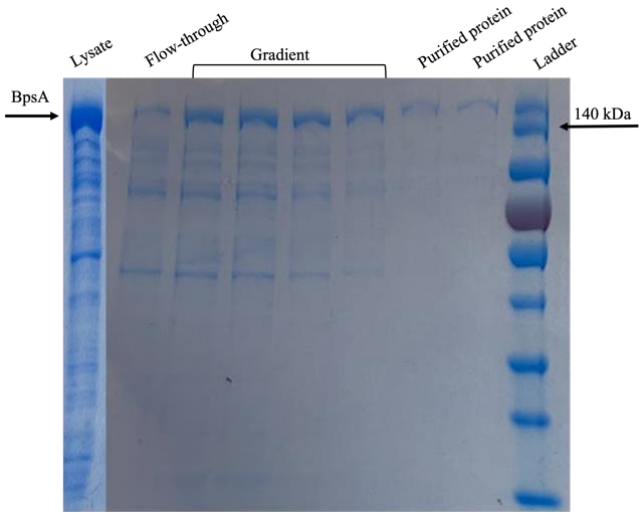
The native coding sequence of BpsA NCBI accession number: BAW81991.1: (continued)	cttcgtcggcggccgcatcagcgacgagcagatcggccacgagggcatgcgcgaggacagcgtccaca tgcgcggtccggccgagctcatccgcgacgacctcgtcagcttcctcccggactacatgatccccaaccgg gtcgtggtcttcgaccggctgccgctgtccgccaacggcaagatcgacgtcaaggcgctcgccgcctccg accaggtcaacgccgagctcgtcgagcgccccttcgtcgccccgcgcacggagacggagaaggagatc gcggcggtctgggagaaggccctgcgccgcgagaacgcctccgtccaggacgacttcttcgagtcgggc ggcaactcgctgatcgccgtcggcctcgtccgcgagctcaacgcgcgcctgggcgtctccctgccgctgc agagcgtcctggagtccccgaccatcgagaagctggcccgccgcctggagcgcgaggtcgcccaggag tcctcgcgcttcgtccgcctgcacgcggagaccggcaaggcccggcccgtgatctgctggccgggtctgg gcggctacccgatgaacctgcgcagcctggccggcgagatcggcctcggccgctcgttctacggcgtcca gtcctacggcatcaacgagggcgagaccccgtacgagaccatcaccgagatggccaagaaggacatcga ggccctcaaggagatccagccggccggcccctacaccctgtggggctactccttcggcgcccgcgtggc cttcgagaccgcctaccagctggagcaggcgggcgagaaggtggacaacctcttcctgatcgccccggg ctccccgaaggtgcgcgccgagaacggcaaggtgtggggccgcgaggcgtccttcgccaaccgcggct acaccacgatcctgttctcggtcttcaccggcaccatttccggtccggacctggaccggtgcctggagaccg tgacggacgaggcctccttcgccgagttcatcagcgagctcaagggaatcgacgtcgacctcgcccgccg gatcatctcggtcgtgggccagacgtacgaattcgagtactccttccacgagctggccgagcgcaccctcc aggcgccgatcagcatcttcaaggccgtgggcgacgactactcgttcctggagaacagcagcggctactc ggccgagccgccgacggtcatcgacctcgacgccgaccactacagcctgctgcgcgaggacatcggcg agctggtgaagcacatccgctacctgctcggcgagtga
E. coli codon-optimized sequence of BpsA part 1.	atgacgttgcaagagacttcggtactggagcctaccttacaggggacaacgaccttacccggcttacttgca cagcgtgtcgctgaacacccagaagctattgcggtggcatatcgtgacgataagttaaccttccgcgagtta gcttcccgcagcgcggcgttggcagattacttagagcatttaggggttagcgcggatgactgtgtaggcctg tttgtagagccatccatcgacctgatggtgggcgcctggggaattttgaatgcaggagcagcttacttacccc tttcgccggagtaccccgaggatcgtcttcgttatatgatcgagaattcggaaacaaaaattattttggcccag caacgtctggttagtcgcctgcgcgagcttgcacctaaggacgttactattgtcacattgcgtgaaagcgaag catttgttcgccctgagggtactgaggcaccagctgcccgtagcgcacgtccagacacccttgcctatgtaat ctatacctctggatctacgggaaagccgaagggtgtgatgattgagcatcgctccattgtaaaccagttagga tggctgcgcgaaacatatgccatcgatcgttccaaggtcatccttcagaaaaccccgatgtcttttgacgcag cgcaatgggagatcctgtcgccggctaacggtgctaccgtcgtaatgggcgccccaggagtgtatgcagat ccggaagggctgattgaaacaatcgtcaagcacaacgtcacaacattacagtgtgtcccaactctgctgcag ggcttaattgatacagagaaattcccggagtgcgtatcactgcagcaaattttctcgggtggagaggccttaa gtcgcctgttagcaattcaaacaacccaggaaatgccgggacgtgcactgattaatgtatatggccccacgg agacaacgatcaatagttctagctttccagtcgatcccgcggatctggatgagggaccccaatcgatctcaat cggatcgcctgtgcatgggacgacttaccatatcttggacaaagagactcttaagcctgtgggagtaggcga aatcggggagttatacattggaggcattcagctggcccgcggctaccttcatcgtgacgacttaactgccga acgctttcttgagattgaactggaggaaggtgccgagccggtacgtttatacaagactggagacttaggcca atggaataatgacgggacggtccagttcgccggtcgtgccgacaatcaagtgaaacttcgtggatatcgtgt ggaattggacgagatctctttagctattgaaaatcacgattgggtgcgtaatgcggctgtaatcgtgaagaatg atgggcgcacaggttttcagaatcttatcgcctgcattgaattatccgagaaggaagccgcccttatggacca gggcaaccacggctcccaccatgcatccaagaaatcgaagttacaagtaaaggctcaactgtctaacccag

Table S2.1. Continued

E. coli codon-optimized sequence of BpsA part 1. (continued)	gactgcgtgacgacgcagaactggcagctcgtcctgcttttgacctggagggcgctgagcctacgccag aacagcgcgcccgtgtatttgcccgcaagacttaccgcttttacgaaggggagccgttacgcaggcgg atttacttgggcttctgggggctaccgttacggcagggtattcgcgtaaagccgcagatttggcgccagcc gaattaggacaaattctgcgttggtttgggcagtacatctcggaagaacgcttgttgccaaagtatgggtat gcatccccgggggctttgtatgctacacagatgtactttgaacttgaaggtgtgggtgggcttaaaccaggg ctactattactatcaaccggtccgtcatcaactggtgcttatttcagagcgtgaggctactggtaaggctacc gctcagattcactttattggcaaaaaatcgggcatcgagcctgtgtacaagaacaacatcctggaggtttta gagatcgaaacgggtcacatggtaggacttttgaacaaattttgcccggcctacggattggacatccacgga tcgtgcttacgaaccagctgtgaaggatttgcttgatgtagctgatgaagactattatcttgggacattcgag ttggtgccacatgccggggcccgcgatgaccaagccgaagtgtatgtccaaacgcatggtggtaaagtg gccggactgcccgaggggcaatatcgttacgaaaatggagaacttactcgcttttctgatgacattgtctta aagaagcatgttattgctatcaaccagtccgtgtaccaggctgcatcg
E. coli codon-optimized sequence of BpsA part 2.	tttggaatttcagtatatagtcgcgcagaggaggaatggctgaaatatattacactgggcaaaaaacttcaa catcttatgatgaacggcttgaacttgggattcatgtccagcggttattcgtccaagactggcaacccccctg ccagcttctcgccgtatggacgctgtattaggcgcgaacggggttgactcggccccaatgtatttttttgta ggaggacgtatctctgatgagcagattggacatgaaggtatgcgtgaagacagcgttcacatgcgcgggg cctgccgaattaatccgtgacgacttggtctccttcctgccagattatatgatcccgaaccgtgtcgttgtgtt cgaccgtttaccgttgagtgctaacggaaaaattgacgtcaaagccttagctgcatcagaccaagtcaac gcagaattggtggagcgcccgttcgtagcgcctcgtaccgagacagagaaagaaattgccgccgtatg ggagaaggctcttcgccgtgaaaatgcctcggttcaagacgacttcttcgaaagtggaggtaattcgctg attgcagtaggattagtccgcgaactgaatgcacgtttaggcgttagccttcccttacagtctgtacttgaga gccctaccattgagaagttggctcgtcgcttggaacgcgaagtcgcgcaggaatcgtcccgctttgtgcg tcttcacgctgaaaccggcaaagcccgtccggtcatctgctggcctggtctgggaggctatccaatgaatt tgcgctcacttgcaggagagatcggcctgggacgctcgttctatggggttcaaagctatggcatcaatga gggggagactccctacgaaaccattactgagatggcgaagaaagacatcgaggccttgaaggagatcc agcccgcagggccatatactctgtgggggtattctttcggcgctcgtgtggcgtttgagactgcatatcag ctggaacaggcgggcgagaaagtggacaacttattcttaatcgccccgggtagtccaaaagtccgtgcg gagaatggcaaagtctggggccgcgaggccagtttcgccaatcgcgggtatacaacgatccttttctccg tctttactgggaccatcagcggcccagatcttgatcgctgcttagagacggtaactgacgaggcatccttc gccgaattcattagtgaactgaaaggcattgacgtggacttggcgcgccgtattattagtgtcgttggccaa acatatgaattcgaatactcgttccatgagttggcggaacgtactttacaggcacctattagtattttcaagg ctgtgggtgacgactattcgttccttgaaaattcgagcggctatagcgctgagccaccgaccgtaattgac ttggacgccgaccactatagcttgttgcgtgaagacatcggagaactggtcaagcatatccgctatttactt ggtgaataa

Table S2.2. Primers used in this study.

Primer name	Sequence
E. coli bpsA part 1 Forward	CACACCAGGTCTCAGCATGACGTTGCAAGAGACTTCGG
E. coli bpsA part 1 Reverse	CACACCAGGTCTCAAACGATGCAGCCTGGTACACGG
E. coli bpsA part 2 Forward	CACACCAGGTCTCACGTTTGGAATTTCAGTATATAGTCGCGC
E. coli bpsA part 2 Reverse	CACACCAGGTCTCAGTGTTATTCACCAAGTAAATAGCGGATATGC
pET28 with *E. coli* bpsA Forward	CACACCAGGTCTCAACACCACCACTGAGATCCGGCTGCTAAC
pET28 with *E. coli* bpsA Reverse	CACACCAGGTCTCAATGCTGCCGCGCGGCACCAG

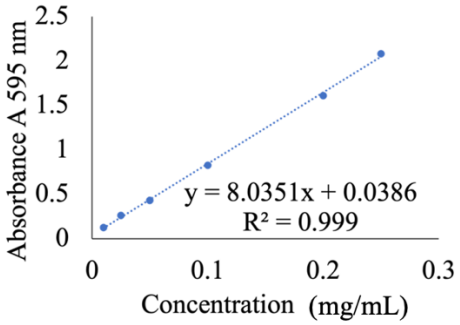


Appendix Figure S2.1. The complete purification process: lysate, gradient (20-60% imidazole), and final purified BpsA.

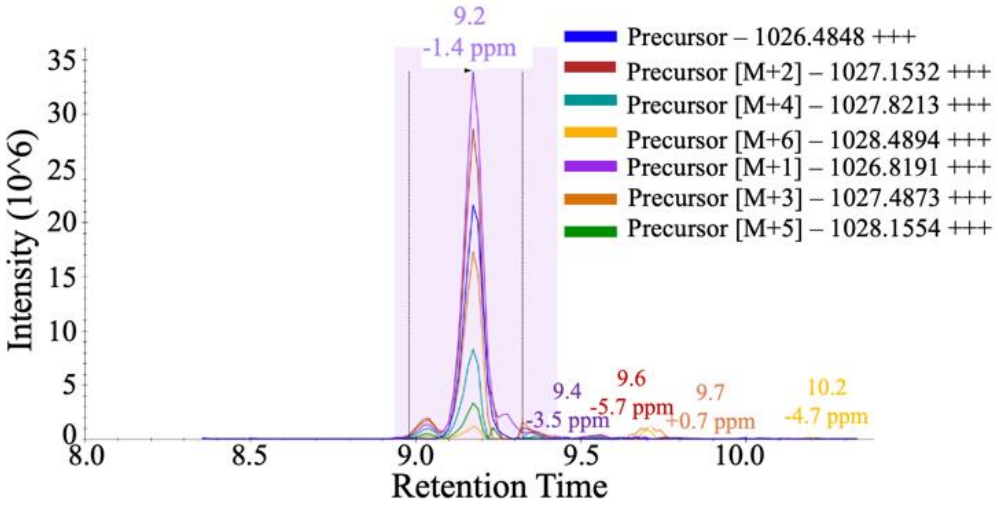
```
alternapyrone           ---IAQRLASLMMIPEADIDAGRPLSAYGVDSLVAVEVRNWVAREMAVEVSVFDVMQN--    55
naphthyridinomycin      EAEIAETFCSLLGVSKVSA--VADFFELGGNSLLVARLTAQLSRTHDVTLPVEQIFRVPT    58
puwainaphycin           EQAIANIVAEVLNIEQVGI--DDNFFELGGNSLNANQVISRLRQTFRVELPLRNLLEEPT    58
N-Bpsa                  EKEIAAVWEKALRRENASV--QDDFFESGGNSLIAVGLVRELNARLGVSLPLQSVLESPT    58
indigoidine             ERRIRDIWQAVLKRDQVSV--TDDFFELGGNSLLAVALVSRLNADFGGAIPLQILFEAPT    58
                           *       :   :..      :   * :** .  :   :       : :  ::.

alternapyrone           ----------     55
naphthyridinomycin      VAGVAAAIEED    69
puwainaphycin           VVELATSMDQ-    68
N-Bpsa                  IEKLARRLER-    68
indigoidine             VERLAAALEAT    69
```

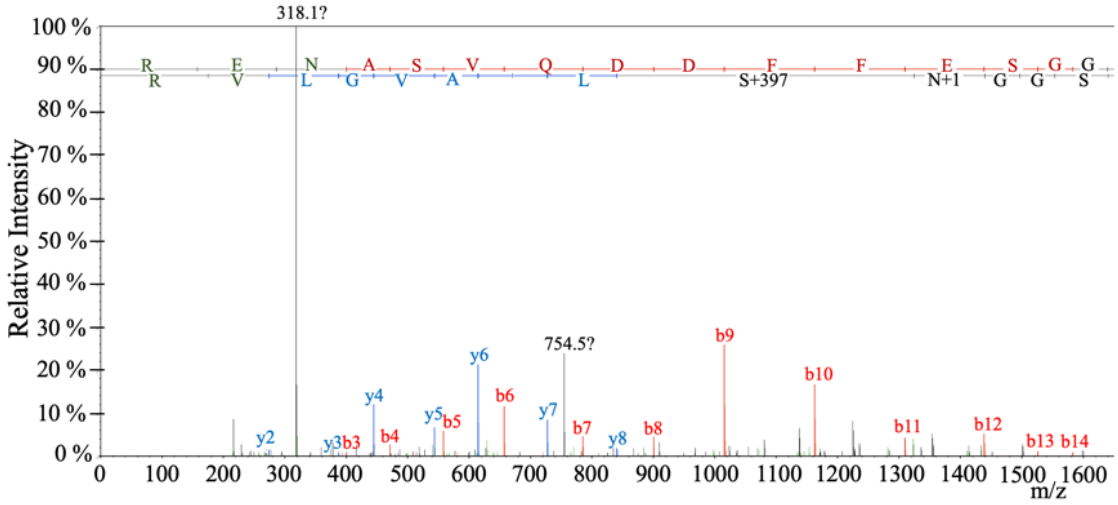
Appendix Figure S2.2. Sequence alignment of the peptidyl carrier protein (PCP) domain between BpsA and four other NRPS.
The green box is the site of conserved serine for phosphopantetheinyl posttranslational modification.



Appendix Figure S2.3. Standard curve of indigoidine.
Different concentration of purified indigoidine was spiked in DMSO, absorbance measurement was at A595 nm.



Appendix Figure S2.4. MS-MS spectra of the holo peptide from BpsA,
ENASVQDDFFESGGN**S**LIAVGLVR with a phosphopantetheiyl modification at the bolded serine.



Appendix Figure S2.5. Fragmentation of the holo peptide from BpsA, ENASVQDDFFESGGNSLIAVGLVR with a phosphopantetheiyl modification at the bolded serine.

CHAPTER 3

EXPRESSION OF BLUE PIGMENT SYNTHETASE A FROM STREPTOMYCES LAVENDUALE REVEALS INSIGHTS ON THE EFFECTS OF REFACTORING BIOSYNTHETIC MEGASYNTHASES FOR HETEROLOGOUS EXPRESSION IN ESCHERICHIA COLI

A version of this chapter was originally published by Tien T Sword, Jaime Lorenzo N. Dinglasan, William J. Barker, Mitchel J. Doktycz, and Constance B. Bailey:

Dinglasan, J. L. N.,* **Sword, T. T.**,* Barker, J. W., Doktycz, M. J., and Bailey, C. B. (2023). Investigating and Optimizing the Lysate-Based Expression of Nonribosomal Peptide Synthetases Using a Reporter System. ACS Synth. Biol. 12, 1447–1460. doi:10.1021/acssynbio.2c00658.

*Authors contributed equally

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

3.1. Abstract

Lysate-based cell-free expression (CFE) systems are accessible platforms for expressing proteins that are difficult to synthesize in vivo, such as nonribosomal peptide synthetases (NRPSs). NRPSs are large (>100 kDa), modular enzyme complexes that synthesize bioactive peptide natural products. This synthetic process is analogous to transcription/translation (TX/TL) in lysates, resulting in potential resource competition between NRPS expression and NRPS activity in cell-free environments. Moreover, CFE conditions depend on the size and structure of the protein. Here, a reporter system for rapidly investigating and optimizing reaction environments for NRPS CFE is described. This strategy is demonstrated in *E. coli* lysate reactions using blue pigment synthetase A (BpsA), a model NRPS, carrying a C-terminal tetracysteine (TC) tag which forms a fluorescent complex with the biarsenical dye, FlAsH. A colorimetric assay was adapted for lysate reactions to detect the blue pigment product, indigoidine, of cell-free expressed BpsA-TC, confirming that the tagged enzyme is catalytically active. An optimized protocol for end point TC/FlAsH complex measurements in reactions enables quick comparisons of full-length BpsA-TC expressed under different reaction conditions, defining unique requirements for NRPS expression that are related to the protein's catalytic activity and size. Importantly, these protein-dependent CFE conditions enable higher indigoidine titer and improve the expression of other monomodular NRPSs. Notably, these conditions differ from those used for the expression of superfolder GFP (sfGFP), a common reporter for optimizing lysate-based CFE systems, indicating the necessity for tailored reporters to optimize expression for specific enzyme classes. The reporter system is anticipated to advance lysate-based CFE systems for complex enzyme synthesis, enabling natural product discovery.

3.2. Introduction

Nonribosomal peptides are a structurally diverse class of natural products with valuable biological activities that include antibiotics, immunosuppressants, insecticides, fungicides, and pigments among others. (145) These short peptides are made independent of the ribosomal machinery, and are instead synthesized by large (> 100 kDa) enzyme complexes called nonribosomal peptide synthetases (NRPSs). (146) NRPSs are modular proteins and can exist as a single module or can be comprised of multiple modules organized into a complex assembly line. Each module has catalytic domains that carry out discrete chemical transformations to build a peptide product from amino acid precursors that are tethered by flexible linkers. (22,30,146) Considerable efforts to reprogram NRPS architectures to generate novel peptides have been made, both in terms of altering selectivity of amino acid recognizing or loading domains through point mutations (147–149) as well as domain swapping. (37,150,151) However, many of these designs are hindered by slow experimental throughput including the ability to screen large enough numbers of constructs to develop adequate design rules. (37,149,152–159) Additionally, there is a desire to study uncharacterized NRPSs identified through genome mining which will enrich the portfolio of valuable nonribosomal peptide structures. (30,160) To enable these endeavors, NRPS gene clusters have been expressed *in vivo* in microbial hosts. (145) However, NRPS yields in cells are compromised by the metabolic burden from expressing such large and complex proteins, the cytotoxicity of their products, as well as the throughput limitations of genetics. (61,62) Hence, there is increasing interest in developing platforms that allow high-throughput profiling of NRPS genes *in vitro* such as in cell-free expression (CFE) systems. (74)

CFE platforms are considered highly suitable for NRPS production because they lack cellular survival objectives, bypassing challenges with *in vivo* expression. (61,74) The open nature of CFE systems is additionally appealing for NRPS characterization and engineering because it can accelerate design-build-test-learn (DBTL) cycles. Direct manipulation of the expression environment affords high-throughput protein synthesis from exogenously added DNA as well as the introduction of non-natural components like non-canonical amino acids, all without requiring extensive cell engineering. (76) To these ends, NRPSs have recently been expressed in both PURExpress (81) and lysate-based (77,78) cell-free systems. The nonribosomal peptide products made in these systems include the pigment indigoidine and rhabdopeptides in the PURExpress format, (81) as well as valinomycin (78) and a diketopiperazine precursor of gramicidin S in extracts. (77) However, CFE is limited by components of the user-defined reaction environment, requiring the definition of effective reagent (e.g., energy substrates, salts, cofactors) concentrations that achieve sufficiently high protein expression yields. (76) (161,162)

Understanding the impact of a set of reaction conditions on CFE is a non-trivial task especially in lysate-based CFE systems which possess complex crosstalk between the

lysate's active transcriptional/translational (TX/TL) and metabolic machineries. (91) For example, components meant to supply protein expression can also be metabolized or used to energize lysate metabolism in ways that unintentionally impact TX/TL processes. (92) This can be especially complicated when synthesizing functional and structurally complex enzymes like NRPSs. The catalytic domains of NRPSs could theoretically utilize reaction components that would otherwise sustain TX/TL, an additional resource competition problem in CFE systems that lacks prior investigation. For example, the NRPS adenylation (A) domain initiates NRP chain elongation by recognizing and activating an amino acid monomer as an adenylate using magnesium-ATP to generate an aminoacyl-AMP intermediate. This domain alone could thus contribute to unintentional lysate metabolism that competes for the TX/TL machinery's synthesis of the enzyme itself, including certain amino acids, the ATP producing energy substrate phosphoenolpyruvate (PEP), and magnesium ions. In addition to these active domains, the larger architectures of these proteins could be responsible for the production of NRPSs and related protein (i.e., polyketide synthases) complexes at low mg/L quantities and with high levels of premature truncation in lysate CFE systems. (77,78)(93) This is presumably because previous efforts have used reaction conditions that are not tailored for full-length translation of these large and complex enzymes, (78)(77,93) but instead were optimized to achieve g/L yields of small and robustly folding proteins such as the reporter proteins chloramphenicol acetyltransferase (CAT) and superfolder GFP (sfGFP). (91)

Different reaction conditions can be sampled to probe limitations to NRPS CFE. Examining this parameter space however can be daunting due to barriers at the level of detection. That is, full-length NRPS CFE has only been validated through SDS-PAGE or mass spectrometric (MS) detection of the product, which by themselves can require time-consuming protocol optimization for different proteins or metabolites, limiting throughput. (77,78) This feedback loop can be shortened with optically detectable peptide tags expressed at the C-termini of the expressed proteins. (71) The use of a terminal tetracysteine (TC) peptide tag that forms a fluorescent complex with an exogenously added organoarsenic molecule has been utilized to monitor *in vitro* protein synthesis in the PURExpress cell-free system. However, its efficiency in the more complex cell extract environment has not been demonstrated. (71) Adapting such reporters for lysate-based systems would have several advantages for NRPS CFE. On top of their relative accessibility and scalability compared to the PURExpress platform, lysate-based systems can be derived from strains that already possess genetic components necessary for post-translationally modifying NRPSs into their functional *holo*-forms. (93,106) Furthermore, alternate lysate systems have been developed for a range of other hosts including *Streptomyces sp.*(80,163–165) and *Bacillus subtilis*, (166) which are natural producers of industrially relevant NRPSs. Adapting reporters to the cell-free lysates would thus allow for applications in a larger range of transcriptional and translational machineries.

Here, the TC tag is used to investigate factors that limit full-length NRPS translation and, consequentially, metabolite production in an *E. coli* lysate-based cell-free system. Specifically, blue pigment synthetase A (BpsA, ~ 140 kDa), a model NRPS, with a terminal TC tag is expressed in *E. coli* lysate reactions. (129,167,168) Based on measurements of BpsA's optically detectable pigment product, the cell-free expressed TC-tagged NRPS is functional in lysates. Furthermore, full-length expression of the tagged protein is detectable in backgrounds supplemented with FlAsH, a fluorogenic biarsenical dye which emits fluorescent signals when bound to the TC tag. (169) The TC/FlAsH reporter system can be used to sample the relative expression of full-length BpsA under different concentrations of PEP, magnesium, and amino acids, and to investigate how the NRPS's catalytic A domain competes against the lysate TX/TL machinery for these resources. Furthermore, conditions defined to improve BpsA synthesis based on TC-FlAsH measurements also increase BpsA activity, as suggested by improved pigment production. Importantly, these conditions also benefit the expression of other NRPSs, tyrocidine synthetase I (TycA) and pyreudione synthetase (Pys). TC tag technology therefore can be leveraged as a strategy for quickly investigating and optimizing the lysate-based CFE of functional, full-length enzyme complexes like an NRPS. (71)

3.3. Methods

3.3.1. Strains and plasmids

E. coli BL21 (DE3) was purchased from VWR (Radnor, Pennsylvania, USA). *E. coli* BAP1 was generously provided by Prof. Christopher Boddy **(**University of Ottawa, Ottawa, Canada) and has been described previously. (106)(170) *E. coli* BL21 Star (DE3) was purchased from New England Biosciences (Ipswich, Massachusetts, USA). A pET28a(+) backbone containing the TC tag sequence, TGTTGCCCGGGTTGCTGT, and 6xHis-tag, CACCACCACCACCACCAC, was designed and ordered from TWIST biosciences (Table S1). The *bpsA* gene from *Streptomyces lavendulae* was codon optimized using Integrated DNA Technology's (IDT) codon optimization tool (https://www.idtdna.com) and was ordered as two gene blocks from IDT (part 1 and part 2) (**Table S3.1**.) These were cloned into the described pET28a(+) vector using the GoldenGate assembly kit (New England Biolabs, part #E1601S) to generate the pET28a(+)-BpsA-TC plasmid. The codon-optimized *bpsA* sequence was mutated with the E315A substitution by mutagenic Gibson assembly (New England Biolabs, part #E2611S) to generate the pET28a(+)-BpsA-TC_E315A construct. The *pys* gene from *Pseudomonas fluorescens* and the *tycA* gene from *Bacillus brevis* were codon optimized and ordered as two gene blocks from IDT (part 1 and part 2) containing the TC tag sequence followed by a StrepII-tag, TGGAGCCACCCGCAGTTCGAAAAA (Table S1). The *pys* and *tycA* gene blocks were cloned into a pET28a(+) vector using the GoldenGate assembly kit (New England Biolabs, part #E1601S) to generate pET28a(+)-Pys-TC and pET28a(+)-TycA-TC

plasmids. The plasmid pJL1-sfGFP was ordered from Addgene (part #69496). Primers used in this study were designed with the J5 software and are listed in **Table S3.2.** (126)

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

E. coli BAP1 was used as a host strain for *in vivo* BpsA-TC expression. Cultures were grown in LB (Miller) broth supplemented with kanamycin at 50 µg/mL. Overnight seed cultures (25 mL) were grown from a single colony at 37°C, shaking at 210 rpm. Flasks (VWR 29171-854) containing 1 L media were inoculated with 10 mL seed culture and incubated at 37°C, shaking at 210 rpm, until an OD_{800} of 0.3-0.4 was reached. OD_{800} rather than the more typical OD_{600} was used to monitor growth to ensure that interference from the signal from indigoidine would not interfere with measurements of cellular turbidity. This strategy was adapted from Beer and coworkers and was reproduced with the correlation between the OD_{600} and OD_{800}. (153) The temperature was then lowered to 18 °C and cooled for 15 minutes prior to induction with 0.5 mM IPTG. The cultures were then incubated for ~ 20 hours at 18 °C prior to harvesting by centrifugation (4000 x g, 45 mins at 4 °C). Cell pellets were resuspended in wash buffer (5mM imidazole, 0.5 M NaCl, 10% v/v glycerol, 50 mM sodium phosphate, pH 7.8) and lysed by sonication. After sonication, the lysate was clarified via centrifugation (11,000 x g, 45 min, 4 °C) and purified via IMAC on an AktaPure system (Cytiva) using a 5 mL HisTrap column (Cytiva). Protein was eluted with elution buffer (400 mM imidazole, 0.5M NaCl, 10% v/v glycerol, and 50 mM sodium phosphate, pH 7.8) using a linear gradient from 20-60% elution buffer for 1 column volume (CV), 1 column volume (CV) of 60-100% elution buffer, and followed with a hold at 100% elution buffer for 3 CV. After purification, the collected fractions were dialyzed with 10K MWCO dialysis tubing (Thermofisher) for 2 passes for a minimum of six hours in storage buffer (50 mM phosphate pH 7.8, 100 mM NaCl, 10% (v/v) glycerol). The protein was then concentrated using a 100K MWCO concentrator (Thermofisher). Purity was assessed via SDS-PAGE using 4-20% pre-made gels (NuSep) (**Appendix Figure S3.2**)

3.3.3. *In vivo* indigioidine production and purification

A 25 mL seed culture of BAP1 harboring the pET28a(+)-BpsA-TC construct was grown overnight (37 °C, 2010 rev/min) in LB medium supplemented with 50 µg/mL kanamycin. After ~ 20 h, 10 mL of seed culture was used to inoculate 1 L media in a flask (VWR 29171-854). Cells were incubated at 37 °C shaking at 210 rev/min. At an OD_{800} of 0.3-0.4, the culture was induced with 0.5 mM IPTG and grown at 16 °C for 20 h then lyophilized for 36 h. The dry cell mass was then washed with three rounds of DI water, methanol, ethanol, isopropanol, and hexanes to remove metabolites, salts, and proteins. Finally, the product was dried for one week under vacuum. Purity was verified via ^{1}H NMR (400 mHz, D_6MSO) δ 11.30ppm (s, NH), 8.18ppm (s, CH), and 6.46ppm (s, NH_2) which is in agreement with the previous literature report. (128)

3.3.4. Cell-free extract preparation

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

3.3.5. CFE reaction preparation

The initial cell-free reaction base mix comprised 1.2 mM ATP, 0.85 mM GTP, 0.85 mM UTP and 0.85 mM CTP; 34.5 µg/mL folinic acid; 0.4 mM nicotinamide adenine dinucleotide (NAD), 0.27 mM coenzyme A (CoA), 4 mM oxalic acid, 1 mM, 1.5 mM spermidine, 57.33 mM HEPES buffer, 10 mM magnesium glutamate, 10 mM ammonium glutamate, 130 mM potassium glutamate, 2 mM each of the 20 amino acids, and 33 mM phosphoenolpyruvate (PEP). Reactions were set up in 10 µL volumes unless otherwise stated. Before condition optimization, reactions were started with 13 ng/µL DNA template and incubated at 30˚C. The type of lysate used and changes to any of these conditions are described in the text. Unless otherwise stated, all CFE reactions were incubated for ~ 20 h. All reactions were incubated in a clear-bottom, black-wall 384-well plate with a non-binding surface (Corning #3544). Whenever possible, reactions were laid out in center wells. Surrounding wells were filled with 1x phosphate buffered saline (PBS) to control the humidity and prevent evaporation. Plates were covered with a hard covered lid and wrapped with parafilm at the sides before incubation.

3.3.6. Indigoidine quantitation in lysates with absorbance measurements

To generate standard curves from pigment absorbance measurements, increasing concentrations of the purified pigment dissolved in DMSO were spiked into BAP1 lysate mock reactions (i.e., reaction mix without DNA). Absorbance measurements were made in a 384-well plate, without a lid, loaded into a BioTek Cytation5 (Agilent) plate reader. The read protocol was set to shake the plate at high speed for 2 s then measure absorbance in selected wells at 590 nm and then at 800 nm. Absorbance values at 590 nm (A590) were divided by values at 800 nm (A800) following a previously described protocol in *E. coli* cells to deconvolute pigment absorbance from the lysate's optical density. (153) To solubilize purified indigoidine, 50 %, 60 %, 70 %, or 80 % DMSO were added to reaction mixes containing purified indigoidine following a protocol previously described in purified BpsA samples. (171) Mixing was performed by gentle swirling using the pipette tip. A590/A800 values taken before DMSO addition were subtracted from values taken after DMSO addition to measure solubilized indigoidine. The resulting values were then normalized to the 0 µM pigment condition. A concentration of 70 % DMSO with this absorbance measurement method could detect a broad range of pigment concentrations (**Figure 3.2B, Appendix Figure S3.2B-D**).

To measure the absorbance of indigoidine produced by cell-free expressed BpsA-TC, base reaction mixes with BAP1 lysate and pET28a(+)-BpsA-TC were incubated at 30˚C. Modifications to the reactions for validating pigment production are described in the text. All reactions were laid out on a 384-well plate and measured after ~ 20 h. A590 and A800 measurements were taken before and after 70 % DMSO addition and normalized as described above. Values reported for the L-Gln, CoA, and DNA experiments were further normalized to appropriate baseline conditions (0 mM L-Gln or CoA and 0 ng/µL DNA).

3.3.7. Comparative quantification of TC-based fluorescence measurements

The distribution of reactions across a plate and subsequent analyses with FlAsH-EDT_2 and ReAsH-EDT_2 labeling reagents (Thermo Fisher #T34563) were evaluated to correct for positional effects from the plate and reader. (172) An appropriate design, tested with protein standards, was then used in all experiments comparing endpoint TC/FlAsH measurements among reactions within a single plate read (e.g., different concentrations of PEP). For each condition tested (e.g., specific PEP concentration), a triplicate set of test reactions (with plasmid) and its corresponding triplicate set of control reactions (without plasmid) were prepared (10 µL volumes). Triplicate test reactions were then plated adjacently on one row (e.g., Row H) and control reactions on another row (e.g., Row I). The test set and control set of each condition were laid out symmetrically to minimize well-to-well variation due to plate design. (172,173) After reaction incubation, 0.5 µL of the labeling reagent was added with gentle swirling of the pipette tip to each well in the

triplicate set of test reactions then immediately to the corresponding control set. This was done for each condition before moving to the next. The plate was then read without a lid on a BioTek Cyation5 (Agilent) reader. For the experiment comparing labeling reagents, fluorescence measurements were taken with top optics every 5 minutes for almost an hour using appropriate filter sets (ReAsH: Exc 590 / 15 nm, Em 610 / 15 nm; FlAsH: Exc 510 / 15 nm, Em 530 /15 nm) and autogain, with 2 s shaking at 567 cycles per minute linear frequency before every read. For every other experiment with FlAsH, measurements were taken after a 10 min incubation period at room temperature (Exc 510 / 15 nm, Em 530 / 15 nm, Gain 70). To isolate signals from TC/FlAsH complexes in the data, signals from control reactions were subtracted from those of test reactions. To mitigate row effects from the detector (which were apparent in our data because our microplate reader reads row-wise) all resulting differences were normalized to the minimum value. (172) Averages with standard error bars were plotted to visualize relative expression changes. For relative intensity analyses (%), the averaged values and standard errors were recalculated by multiplication to a constant of 100 over the maximum average value.

3.3.8. BpsA concentration estimation using TC/FlAsH measurements

Triplicate reactions containing pET28a(+)-BpsA-TC and triplicate controls with no DNA were set up on the same row in a microplate with reaction mixes (without DNA) laid out on an adjacent row. All reactions were prepared in 10 µL volumes. After a ~ 20 h incubation at 25 ˚C, the blank reaction mixes on one row were spiked with 1 µL of water or purified BpsA-TC protein dissolved in water to prepare triplicate sets of reactions containing 0, 50, 100, 200, 400 µg/mL protein for standard curve generation. The test and control reactions on the adjacent row were spiked with 1 µL water to equalize volumes. FlAsH (0.5 µL) was added to each well and the plate was read following the protocol described above. Signals for reactions from either row were normalized to the minimal signal within the same row (i.e., lowest signal among the 3 control reactions). The appropriate dilution factor was applied. Linear regression using average normalized values from standard reactions was used to estimate the concentration of expressed BpsA-TC in the test reaction. This analysis was performed for each reaction mix (base vs optimized) when quantifying BpsA-TC since TC/FlAsH complex formation is likely to alter in different reaction backgrounds (**Appendix Figure S3.5**).

3.3.9. SDS-PAGE gel and Western Blot analyses

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

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

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

3.3.10. TycA and Pys concentration estimation

TC/FlAsH measurements from quadruplicate reactions initiated with pET28a(+)-Tyc-TC and pET28a(+)-Pys-TC and corresponding no DNA controls in base and optimized conditions were used to estimate concentrations on standard curves. Separate standard curves for standard and optimized CFE conditions were prepared with purified BpsA-TC as described above. Pooled lysate reactions were loaded onto 7.5% SDS-PAGE gels as described above. For StrepII-tag elutions of CFE reactions, quadruplicate reactions were set up similarly and pooled the next day. Approximately 45 µL of pooled reactions were run through Strep-Tactin®XT Spin Columns (IBA Life Sciences, #2-4150-025). Columns were washed 4x and eluted with 50 µL elution buffer following the manufacturer's protocol. Elutions were loaded onto SDS-PAGE gels to visualize purified proteins.

3.3.11. Quantification of active sfGFP

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

3.4. Results and Discussion

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

The activities of enzymes can affect their cell-free production if the enzymes' catalytic domains share substrates with the TX/TL machinery. To study this phenomenon

for NRPS CFE using TC tag technology, it is important that the TC tag itself does not abolish the catalytic function of cell-free expressed NRPSs in *E. coli* lysates. To this end, a pET28a(+) vector containing the BpsA coding sequence and a C-terminal TC tag sequence to express TC-tagged BpsA (BpsA-TC) was constructed. BpsA is a great model for investigating NRPS CFE due to its relatively simple structure as a single module NRPS and the optical detectability of its product. (129,167,168) BpsA synthesizes one molecule of indigoidine, a blue pigment that absorbs light at a wavelength of 590 nm, from two molecules of L-glutamine (L-Gln) (**Figure 3.1A**). (171) To do so, BpsA must be functionally activated into its *holo*-form by a 4′-phosphopantetheinyl transferase (PPTase) which phosphopantetheinylates the PCP domain. BAP1 is a BL21(DE3) derivative strain with a genomically integrated *sfp* gene that expresses a PPTase. (106)

Detecting the blue pigment in BAP1 lysate reactions incubated with the plasmid construct would indicate successful expression of full-length and functional BpsA-TC protein. An assay was therefore established, allowing indigoidine measurement in a lysate-based cell-free system. To achieve this, purified indigoidine was spiked into lysate reactions without DNA. A sufficiently broad range of pigment concentrations could be detected by normalizing the absorbance of the pigment at 590 nm to lysate turbidity absorbance at 800 nm, and then coupling these normalized values to the addition of 70 % DMSO directly to the mock reactions (**Figure 3.1B, Appendix Figure S3.1**). This optimized protocol was adapted from previous reports of indigoidine measurement in cell cultures and purified samples. (153,171)

With an assay established, pigment production from the pET28a(+)-BpsA-TC construct could be tested in overnight CFE reactions. Reactions initiated with higher concentrations of the indigoidine precursor, L-Gln, have increasing normalized absorbance values, implying pigment production (**Figure 3.1C**). Separately, the base reaction mix was supplemented with CoA, from which the PPant post-translational modification that functionalizes the NRPS PCP domain is derived. Absorbance measurements also increase when up to 2 mM CoA was added to the reaction (**Figure 3.1D**). The improvement in pigment production in CoA-supplemented reactions is drastic even without supplemented L-Gln. To further validate that the measurements correspond to functional BpsA expressed from construct, 1 mM CoA was added to base reaction mixtures containing varying concentrations of the plasmid. Consistent with previous observations, the DNA template concentration affects protein expression and therefore product formation (**Figure 3.1E**). Altogether, the system expresses BpsA-TC and the terminal TC tag does not abolish BpsA folding and activity. Additionally, while optimizing enzyme turnover is outside the scope of this work, it is important to note that these lysates generated a maximum concentration of ~ 900 µM or ~ 223 mg/L indigoidine, which is higher than indigoidine titers reported in the PURE cell-free system (250 µM). (81)

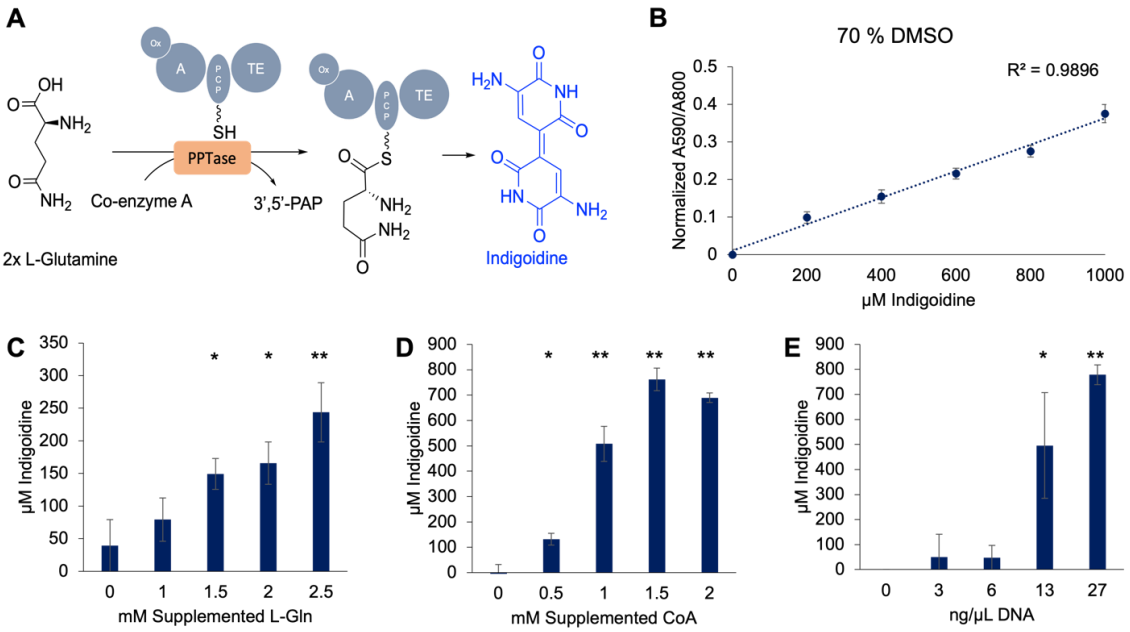


Figure 3.1. Indigoidine production via BpsA in CFE reactions.
A) Schematic of BpsA-catalysed conversion of 2x L-Gln to 1x indigoidine. The *apo*-form of BpsA is activated by the addition of a coenzyme A and the presence of PPTase. The active *holo*-form BpsA converts two L-Gln molecules to indigoidine, involving oxidation (Ox), adenylation (A), peptide carrier protein (PCP), and thioesterase (TE) domains. **B)** Standard curve of indigoidine generated by spiking increasing concentrations of purified indigoidine into BAP1 lysate CFE mock reactions. Indigoidine in the lysate background was solubilized by adding 70 % DMSO directly to individual reactions. Absorbance measurements were taken at 590 nm (A590) and divided by signals at 800 nm (A800) to deconvolute pigment absorbance from lysate turbidity. (153) Afterwards, A590/A800 signals before DMSO addition were then subtracted from measurements after 70 % DMSO addition to measure soluble pigment. (167) For visualization of trends, values were averaged, compared to a baseline condition, and plotted with error bars representing the standard error of the mean (n=3). **C, D, E)** Validation of indigoidine production from functional BpsA-TC protein in CFE reactions using the indigoidine detection method established here. CFE reactions with 13 ng/µL plasmid were initiated with increasing concentrations of **C)** supplemented L-Gln (no CoA) or **D)** supplemented CoA (no L-Gln). **E)** CFE reactions containing 1 mM CoA were initiated with different concentrations of plasmid DNA. Asterisks represent significant differences compared to the no L-Gln, no CoA, or no DNA conditions, evaluated with a Student's t-test. * p-value < 0.05, ** p-value < 0.01.

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

Given that a functional TC-tagged NRPS can be expressed in the system, a protocol for TC tag detection in lysate-based CFE reactions could be established. The TC tag, optimized for *in vivo* applications, has successfully been adapted for an *in vitro* expression study in the PURExpress system, where the tag was complexed with the organoarsenic ReAsH reagent to emit a fluorescent signal at 610 nm. (71) Reportedly, the TC/ReAsH complex has a low signal-to-noise ratio in lysate reaction backgrounds, but the use of controls to mitigate noise has not been explicitly stated in the literature. (71) Alternatively, the organoarsenic FlAsH reagent, which emits at 530 nm, has been used to detect TC-tagged proteins with higher specificity than ReAsH in cells but has not been tested in lysate reactions. (169)

In overnight reactions that expressed BpsA-TC, signals above background are highest after a 10-minute incubation period with the same concentration of either reagent (**Figure 3.2A**). This is sufficient time for the TC/fluorophore complex to form based on a previously described protocol in cells. (169) The signal in BpsA-TC expressing reactions however decreases at later timepoints, consistent with the manufacturer's descriptions of these reagents binding to the background. The FlAsH reagent was therefore selected over ReAsH to obtain end-point measurements of BpsA-TC production since the former emitted less noisy and variable signals. To verify that the signal obtained at 10 minutes corresponds to synthesized BpsA-TC, FlAsH was added to overnight BAP1 lysate-based reactions prepared with varying DNA concentrations. Evidently, increasing concentrations of the expression template in reactions results in higher fluorescence readouts, suggesting the detectability of BpsA-TC in lysate reactions with the FlAsH reagent (**Figure 3.2B**). However, the signal does not increase between the 6 and 13 ng/µL DNA conditions which contradicts the indigoidine production trend in the same lysate (**Figure 3.1E**). It was hypothesized that fluorescence signals from reactions that produce significant pigment quench due to possible overlap in the indigoidine absorbance and TC/FlAsH emission spectra. To evaluate this possibility, *apo*-form BpsA-TC was purified from *E. coli* BL21(DE3) cells lacking the *sfp* gene for PPTase expression. Purified BpsA-TC was then added into BL21(DE3) lysate-based reaction mixes with or without purified indigoidine, demonstrating that the presence of pigment does dampen TC/FlAsH emissions (**Figure 3.2C**). This quenching issue should thus be considered when pairing the fluorescence protein labeling method with the detection of other enzymatic products with optical properties. Moving forward, the TC/FlAsH method is further optimized in lysates from the *E. coli* BL21(DE3) strain, which lack the PPTase and do not produce pigment after overnight BpsA-TC expression (**Figure 3.2D**), and in its derivative BL21 Star(DE3) strain, which should similarly express the *apo*-form of the enzyme**.**

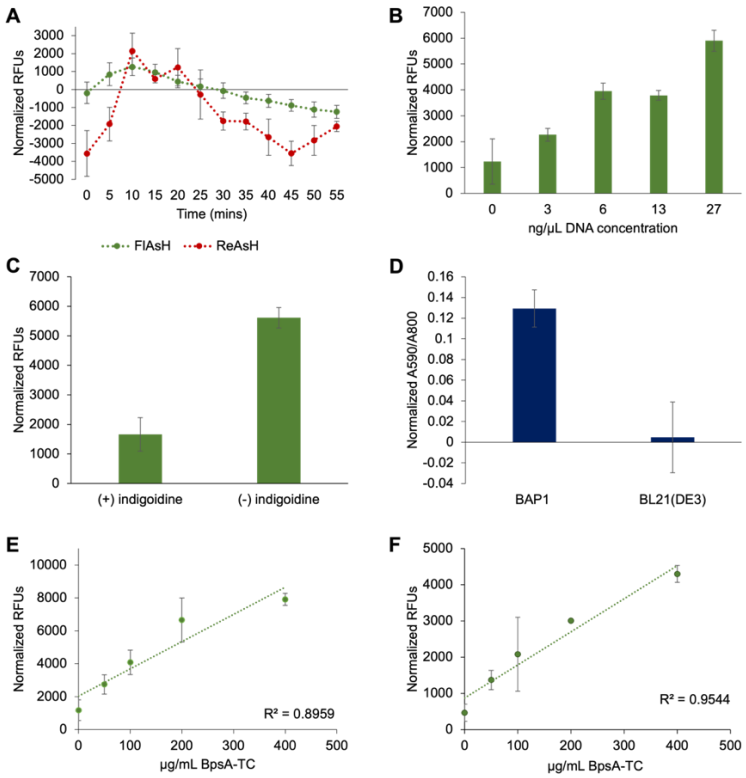


Figure 3.2. Optimization of TC/FlAsH complex measurements in E. coli lysate-based CFE reactions.

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

To use TC/FlAsH signals for measuring protein CFE in lysate reactions, an optimal concentration of FlAsH for detecting a range of BpsA-TC concentrations in *E. coli* lysates must be established. To this end, BpsA-TC expressed and purified *in vivo* was used to generate standards for optimization of the TC/FlAsH method (**Appendix Figure S3.2**). Evidently, fluorescence signals for a range of protein concentrations (0 to 400 µg/mL) in BL21(DE3) lysates are most distinguishable with 30 µM FlAsH reagent compared against concentrations of 7.5 and 15 mM FlAsH, seemingly because these lower concentrations are insufficient to overcome background noise from the lysate **(Figure 3.2E, Appendix Figure S3.3)**. 30 µM FlAsH is similarly useful in BL21 Star(DE3) lysate reactions (**Figure 3.2F**). Thus, TC/FlAsH measurements can be obtained when either lysate is used as an expression platform for BpsA. However, BpsA-TC can be best validated via SDS-PAGE when expressed in BL21 Star(DE3) lysate reactions incubated at 25 ˚C (**Appendix Figure S3.4**). Since SDS-PAGE can serve as an orthogonal method for validating correlations of TC/FlAsH signal trends to protein CFE, succeeding reactions were performed in these lysates at this temperature.

Having established parameters for detecting BpsA-TC in lysates using FlAsH-based and orthogonal SDS-PAGE methods, the TC/FlAsH complex's utility for measuring NRPS CFE under different reaction conditions could then be evaluated. The large sizes, complex structures, catalytically active domains, and precursor (i.e., amino acids) and cofactor requirements (i.e., magnesium and ATP for adenylation) of NRPSs could constrain TX/TL processes and therefore their full-length synthesis, but evidence of such limitations and methods for probing these with high throughput have not been reported. (93) To address this, the ability of the optimized TC/FlAsH detection protocol to interrogate NRPS CFE is assessed using BpsA with an active A domain (a core component of NRPSs) as a model. (30) This strategy is specifically applied here to determine whether resource competition problems exist for three crucial CFE reaction components that are known to energize TX/TL and background lysate metabolism but could also feed into the typical NRPS A domains—PEP, magnesium, and amino acids. However, a potential caveat of the current fluorescence labeling method is the influence of the physicochemical environment on TC/FlAsH complex formation. Indeed, drastically changing the composition of reactions by removing all or one of these three components affects TC/FlAsH signals (**Appendix Figure S3.5**). A possible explanation for this is that altering the ionic strength (by removing magnesium) and pH (by removing amino acids or PEP) of the system impacts binding of fluorophore's arsenic groups to the tag's cysteine groups. For this reason, the effect of each reagent on BpsA-TC CFE was evaluated separately and optimized sequentially. SDS-PAGE gels were crucial for validating a testable range of reaction concentrations wherein trends in fluorescence measurements correlate with TC-tagged protein. The results show that despite the method's caveats, expression trends can be determined from TC/FlAsH measurements for the reported ranges of reagent concentrations that are still relevant to lysate-based CFE.

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

PEP

An obvious potential limitation to functional NRPS CFE is ATP availability. TX/TL processes that depend on ATP include dNTP synthesis, aminoacyl-tRNA charging, and GTP regeneration for peptide bond formation. (91) In the PANOx-SP system, ATP generation is sustained by ADP phosphorylation when the energy substrate PEP is converted to pyruvate. Without other limiting factors, a higher PEP requirement should theoretically better fuel the sufficient expression of a large protein (> 100 kDa) compared to smaller model CFE proteins (e.g., CAT and sfGFP) which are < 30 kDa. (91) The degradation of PEP itself however contributes to inorganic phosphate accumulation which in turn inhibits cell-free protein synthesis. (174) Furthermore, ATP overproduction negatively impacts CFE efficiency. (92) Thus, for any given protein, an optimal range of initial reaction PEP concentrations exists where PEP energizes but is not inhibitory towards TX/TL. (174–176) This range is further defined by central metabolic reactions that consume ATP in an unproductive manner (i.e., not contributing to protein expression). (92,177,178) In the case of NRPS CFE, the A domain that is characteristic of these enzymes could also deplete the ATP pool, further contributing to expression limitations. (30,160) The TC/FlAsH protocol was thus tested as a potential strategy for assessing the sensitivity of BpsA expression to PEP concentrations, and for investigating whether the catalytic A domain also competes for this TX/TL resource.

Fluorescence reads of the BpsA-TC reactions supplemented with FlAsH reagent indicate a trend in signals that were corroborated by an SDS-PAGE analysis, demonstrating the method's applicability for measuring relative expression in reactions fed different concentrations of PEP. To interrogate the interdependency between NRPS TX/TL and NRPS catalytic activity, a construct that would express BpsA-TC with a glutamine-to-alanine substitution at a catalytically essential residue in the active site of the A domain was generated. This point mutation (E315A) disrupts magnesium ion binding which stabilizes ATP in the active site (**Appendix Figure S3.6**). (179) The mutant protein, BpsA-TC_E315A, would thus have an inactive A domain due to disrupted ATP binding. Where BpsA-TC synthesis in this system peaks with higher starting PEP concentrations (55-77 mM), the maximal BpsA-TC_E315A expression already occurs in the reaction initiated with 33 mM PEP (**Figures 3.3A & B**). Trends for BpsA-TC expression under the tested PEP concentrations are corroborated by SDS-PAGE gels (**Appendix Figure S3.7A**). These results confirm that functional NRPS TX/TL can have higher energy substrate requirements than proteins of the same size because the catalytically active A domain contributes to the additional catabolism of ATP. Indeed, BpsA-TC_E315A is expressed more than BpsA-TC in reactions that are read at the same time and that were initiated with the same amount of PEP.

Reactions were also set up with sfGFP to compare the BpsA trends to those of a smaller, robustly folding protein (**Figure 3.3C**). Interestingly, relative expression patterns for sfGFP resemble those of BpsA-TC. However, sfGFP yields significantly increase as initial PEP concentrations rise to 77 mM while BpsA-TC_E315A expression has no significant change beyond the 33 mM PEP condition. Both BpsA-TC_E315A and sfGFP should not compete with TX/TL processes for ATP, so the disparity in these trends may be attributed to other factors that limit the full-length synthesis of a larger and likely less soluble protein than sfGFP.

Unexpectedly, BpsA-TC synthesis in reactions initiated with 100 mM PEP is substantially reduced which may also be attributed to the active A domain (**Figure 3.3A**). In the conventional PANOx-SP system, PEP concentrations beyond the optimum cause high inorganic phosphate accumulation due to the degradation of PEP itself. (174) In the current system, this phenomenon may be partially responsible for the significant inhibition of BpsA-TC synthesis in the 100 mM PEP reactions compared to all other conditions. However, BpsA-TC_E315A yields in reactions started with 100 mM PEP are not as drastically reduced (**Figure 3.3B**). These data imply that the A domain contributes to the inhibition of NRPS TX/TL when PEP is increasingly available, hypothetically because there is a larger ATP pool to liberate pyrophosphates from during amino acid adenylation. (30,174) To test whether the active A domain indeed leads to lower expression of BpsA when more ATP is available, reactions containing constructs for BpsA-TC or BpsA-TC_E315A were fed 77 mM PEP and increasing supplemental concentrations of ATP. PEP was still added to the system because adding ATP alone resulted in no protein expression. Compared to reactions without additional ATP, increasing ATP concentrations only reduces BpsA-TC expression (**Figure 3.3D**, **Appendix Figure S3.7B**) but improves or has no significant effect on BpsA-TC_E315A synthesis (**Figure 3.3E**). However, BpsA-TC_E315A synthesis decreases when reactions are supplemented with inorganic phosphates (**Figure 3.3F**). Altogether these data imply that the active A domain's liberation of free phosphates is increasingly deleterious to expression when higher ATP pools are available.

Magnesium

Magnesium ions are essential in almost all stages of TX/TL. Among a multitude of roles, they stabilize mRNA molecules and ribosomes, are required in the aminoacylation of tRNAs, regulate the ionic strength of the lysate reaction environment, and serve as a co-factor for several TX/TL relevant enzymes. (180,181) Particularly in cell extracts, where magnesium is also sequestered by enzymes of central metabolism, the magnesium

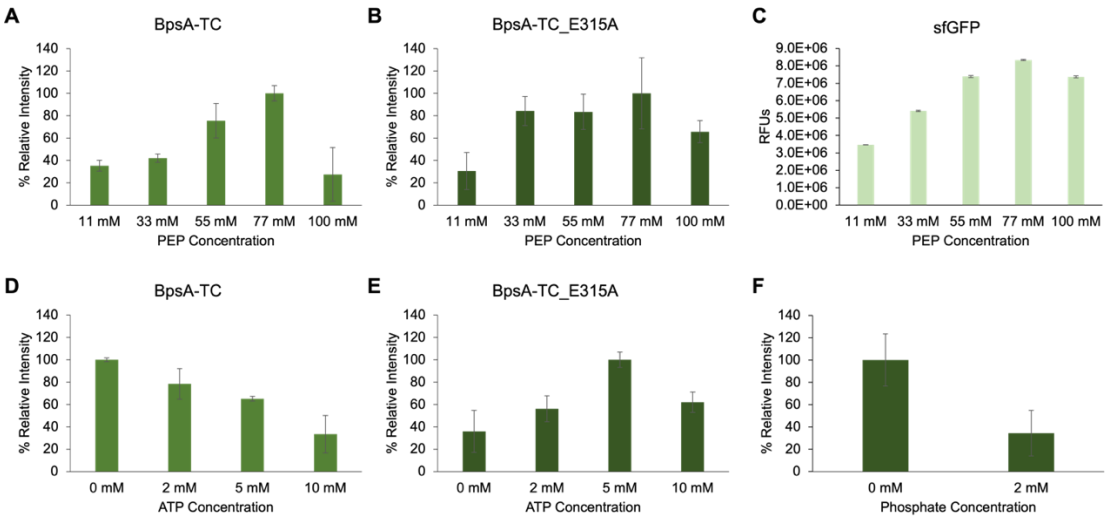


Figure 3.3. TC/FlAsH analysis of energy substrate requirements in NRPS CFE using BpsA as a model.

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

requirement for maximum synthesis can vary per protein and even per batch of lysate. (91,174,182) Methods for routinely optimizing the magnesium concentration of a specific target protein in a lysate-based cell-free system are thus necessary. Such protocols would be particularly beneficial in NRPS CFE efforts, given that enzymes belonging to this structurally diverse family may have varying magnesium concentration optima for TX/TL based on size differences alone. (174,182) Additionally, NRPSs themselves use magnesium at their A domains to stabilize ATP binding when adenylating an amino acid substrate. (30,160,179) Thus, along with metabolic processes that are counterproductive to TX/TL, this domain could also affect magnesium homeostasis in NRPS CFE reactions if it were to sequester free magnesium during *in vitro* synthesis.

If applicable, the TC tag and FlAsH fluorophore can be leveraged to understand and meet the magnesium requirements for the functional expression of an NRPS. BpsA-TC reactions were therefore prepared with varying concentrations of magnesium and 77 mM PEP since this concentration allowed the highest BpsA expression. The results, confirmed by SDS-PAGE, show that an optimal magnesium concentration for expression of the wild-type enzyme can be defined by the approach (**Figure 3.4A, Appendix Figure S3.8**). The expression trends for these reactions differ from those of BpsA-TC_E315A and sfGFP in the same reaction backgrounds, showing that the active A domain impacts NRPS CFE magnesium requirements and reinforcing that the optimal concentration of this TX/TL resource can vary from protein to protein (**Figures 3.4B & C**). (182) Compared to sfGFP and BpsA-TC_E315A, BpsA-TC requires more magnesium for peak expression in the current system. A likely explanation is that additional phosphates liberated by the active A domain sequester free magnesium. (174)

Amino acids

Amino acids are the building blocks of peptides and are thus added into CFE reaction mixes to supply protein synthesis. (183) Like PEP and magnesium, this reaction component has been optimized for CFE in the *E. coli* PANOx-SP system, which is commonly initially supplemented with 2 mM of the 20 essential amino acids. (183–185) This amount reportedly increases yields of CAT compared to lower or higher concentrations and has been previously adapted for NRPS expression in *E. coli* lysates. (77,78,184) However, the sensitivity of NRPS expression to a range of amino acid concentrations can differ from a protein like CAT for a number of reasons, and should be probed with TC/FlAsH measurements. As an NRPS is synthesized, its A domain may recognize and adenylate certain amino acids that would otherwise be used for TX/TL. (30,155) Additionally, NRPSs can have sequences with diverse amino acid distributions or A domains with different amino acid specificities. (149) Also, compared to sfGFP, the CFE of a much larger protein such as an NRPS would hypothetically require more free amino acids.

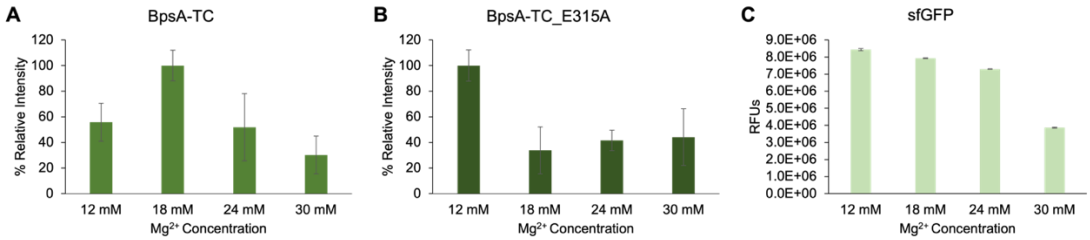


Figure 3.4. TC/FlAsH analysis of magnesium requirements in NRPS CFE using BpsA as a model.

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

The initial concentrations of the amino acid mix (containing all 20 essential amino acids) in BpsA-TC expressing reactions were varied. Reactions were prepared with PEP and magnesium concentrations optimized for BpsA expression thus far (77 mM PEP and 18 mM magnesium). TC/FlAsH measurements from these reactions demonstrate no significant differences in expression levels between the 2, 4, and 6 mM conditions (**Figure 3.5A**). However, a substantial decline occurs when 8 mM of the amino acid mixture is used. These results are consistent with SDS-PAGE data (**Appendix Figure S3.9A**). On the other hand, BpsA-TC_E315A synthesis shows no significant change across all conditions, implicating the A domain's role in impairing BpsA expression presumably because of its interaction with L-Gln (**Figure 3.5B**). (123) To investigate this, reactions containing 2 mM total amino acids and the constructs for BpsA-TC or BpsA-TC_E315A were also initiated with increasing concentrations of L-Gln. However, extra L-Gln in the reaction impairs expression of both proteins, but the mutant BpsA has a more gradual decline with increasing L-Gln supplementation (**Figure 3.5D & E, Appendix Figure S3.9B**). Altogether, these findings are contrary to the assumption that reactions expressing BpsA significant differences in expression levels between the 2, 4, and 6 mM conditions (**Figure 3** with A domain mediated amino acid adenylating activity would benefit from more amino acids than the mutated counterpart. Expression trends for BpsA-TC and BpsA-TC_E315A also contradict the expectation that yields of larger proteins will increase with amino acid concentrations. Where sfGFP synthesis peaks at 2 mM, which was expected of this small protein in the PANOx-SP system, amino acid supplementation has no positive effect on the expression of either the wild-type or mutant BpsA (**Figures 3.5A-C**). Amino acid oxidation, which has been observed during lysate-based CFE, could be partially responsible for these observations. (174) Reactions that actively deplete amino acids like glutamate form phosphates. (186) Amino acids also serve as secondary energy substrates that can drive ATP generation, (174) which in turn can also release phosphates. The deleterious effect of the free phosphate pool on BpsA expression is exacerbated by the A domain's participation in phosphate liberation (**Figures 3.3C-E**), which may explain the higher sensitivity of wild-type BpsA to amino acid concentrations compared to its mutant counterpart (**Figures 3.5A & B**). Further, the oxidation of certain amino acids in lysates reportedly occurs quickly, which may prevent the prolonged translation of larger enzymes that require these amino acids in batch reactions. For example, compared to one mole of sfGFP, one mole of BpsA requires 9x and 11x more moles of arginine and tryptophan, respectively, and both of these amino acids are depleted within 30 minutes in lysates.(174,186)

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

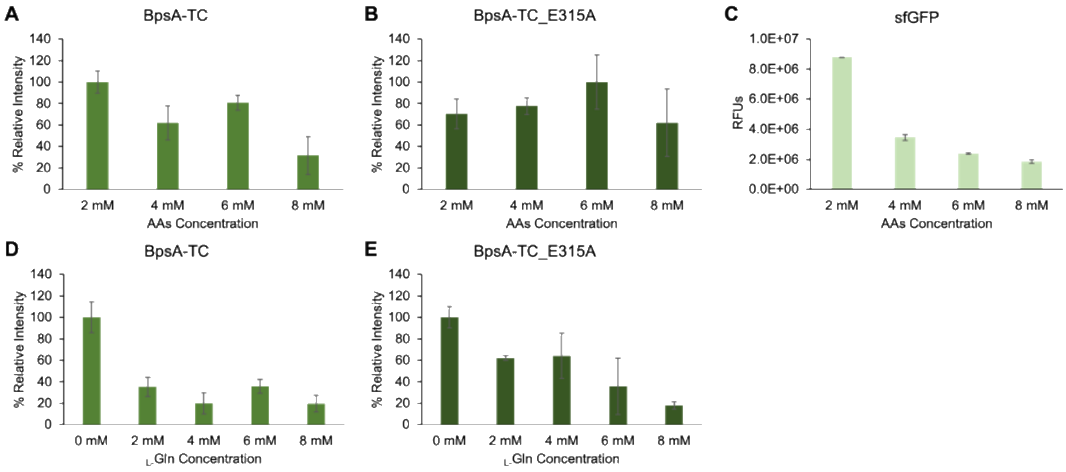


Figure 3.5. TC/FlAsH analysis of amino acid (AA) requirements in NRPS CFE using BpsA as a model.

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

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

3.4.4. Metabolite production improves with optimized NRPS expression conditions

NRPSs are valuable for their bioactive peptide products. The previous sets of experiments show that full-length NRPS expression can be optimized by testing different CFE conditions with a C-terminal TC tag and the FlAsH dye. Higher concentrations of the fully synthesized enzyme should enable higher nonribosomal peptide production as well. To assess this, both BpsA-TC and indigoidine concentrations were estimated and compared in reactions made with base (containing 33 mM PEP and 12 mM magnesium) and optimized (containing 77 mM PEP and 18 mM magnesium) reaction mixes. Using TC/FlAsH measurements and appropriate standard curves, the average concentrations of BpsA-TC in reactions prepared with the optimized mix and base mix are estimated to be 95 µg/mL and 56 µg/mL, respectively (**Figure 3.6A**). Reactions expressing BpsA-TC prepared with either mix were additionally initiated with purified PPTase, CoA, and L-Gln then incubated overnight. Using the colorimetric assay, titers of indigoidine were calculated from a standard curve and show a similar improvement after PEP and magnesium concentration optimization (160 µM and 99 µM indigoidine in the optimized and base reactions, respectively) (**Figure 3.6B**). As expected, optimizing CFE facilitates increased nonribosomal peptide production. Notably, the indigoidine titer in the optimized BL21 Star(DE3) lysate is lower than the maximum concentration obtained in the BL21(DE3)-derived BAP1 lysate (~ 900 µM) (**Figure 3.1D**), despite BL21(DE3) being less productive for NRPS CFE compared to BL21 (Star)DE3 (**Appendix Figure 3.4A**). (77) PPTase, CoA, and L-Gln concentrations can therefore be optimized to enable higher NRPS turnover in BL21 Star(DE3) lysates. Alternatively, a BL21 (Star)DE3 strain with a genome-integrated *sfp* gene could benefit the CFE of NRPSs, especially those that catalyze high-value but cytotoxic metabolites.

3.4.5. Conditions optimized for BpsA allow or improve the CFE of other NRPSs.

Standard CFE conditions commonly used in previous reports of lysate-based NRPS or NRPS-like protein expression include the preparation of reactions using the described base mix and their incubation at 30 ˚C. (77,78,93) In theory, the CFE of NRPSs with

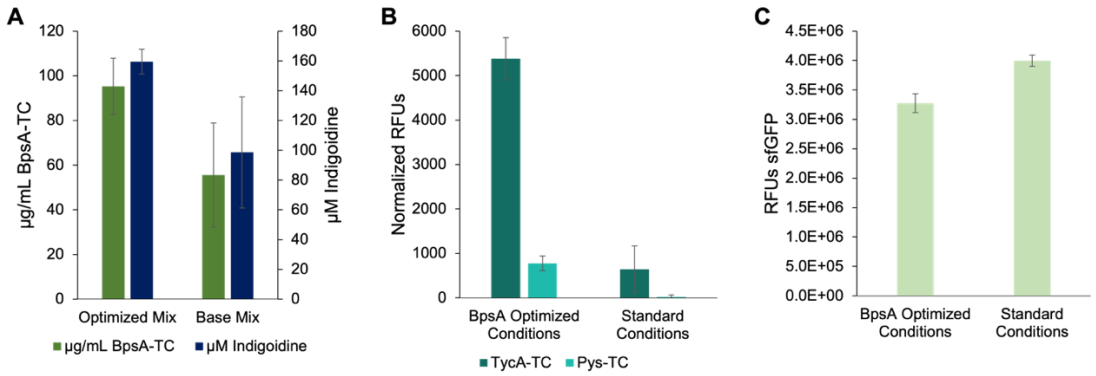


Figure 3.6. Conditions optimized for BpsA improve indigoidine production and the expression of similar NRPSs in lysates.

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

similar architectures to BpsA will benefit from the total set of reaction conditions optimized here (i.e., optimized mix and 25 ˚C incubation) due to similar size and cofactor requirements. Other known monomodular NRPSs include tyrocidine synthetase I, TycA (124 kDa), which incorporates and isomerizes an L-phenylalanine residue into the cyclic antibiotic decapeptide tyrocidine, (187) and pyreudione synthetase, Pys (142 kDa), which adenylates L-proline in the synthesis of amoebicidal pyreudiones. (188) In addition to being close in size to BpsA, these NRPSs each have an active A domain. TC- and StrepII-tagged versions of these NRPSs were expressed using standard CFE conditions and conditions optimized here for BpsA. TC/FlAsH measurements for both TycA-TC and Pys-TC show that these proteins are better synthesized under conditions optimized for BpsA expression (**Figure 3.6B**). Concentrations of these proteins in lysate reactions were estimated using BpsA-TC/FlAsH standard curves prepared under either optimized or standard conditions. TycA-TC expresses particularly well in the optimized system (~ 385 μg/mL) and can be visualized by a thick band in an SDS-PAGE gel of pooled lysate reactions (**Appendix Figure S3.10A & B**). Pys-TC has low expression even in optimized conditions (~ 38 μg/mL) and could not be observed in gels of lysate reactions (**Appendix Figure S3.10B**) or StrepII-tag elutions and thus required the greater sensitivity provided by the fluorescent TC tag to detect (**Appendix Figure S3.10A**). Under standard conditions, neither proteinseems to be significantly expressed (**Appendix Figure S3.10A & B**). On the other hand, the standard condition set, which was initially optimized for sfGFP, (176) is still more beneficial for the synthesis of this protein compared to conditions optimized here for BpsA (**Figure 3.6C**). (176) This implies that sfGFP is not universally informative when tailoring extracts for CFE, and further reinforces the need for reporters like the TC/FlAsH system when developing lysate-based platforms for different types of proteins.

Together, these results confirm that CFE conditions optimized for a particular enzyme allow or improve the synthesis of proteins with similar architectures and cofactor requirements. Thus, while conditions optimized for sfGFP production may not be universally optimal for larger biosynthetic enzymes, generalizable reporter tools like TC/FlAsH may be leveraged to construct model systems for studying and defining better CFE conditions of whole enzyme classes with high throughput. Notably, discrepancies in how much these conditions improved TycA vs Pys or BpsA expression suggest that TC/FlAsH measurements will be useful for further probing and optimizing CFE on a protein-by-protein basis.

3.5. Conclusions

Lysate-based CFE is an accessible and scalable approach for expediting the expression of proteins that are difficult to synthesize *in* vivo, such as natural product forming NRPSs. (74,76,89) Lysates may additionally be derived from strains that can post-translationally modify NRPSs into their active *holo*-forms, (93,106) or microbial species that are already heavy producers of NRPSs. (80,163,165,166) However, lysates possess

complex interactions between the TX/TL machinery and metabolic reactions in the lysate that are unproductive to protein expression. (92) The latter may include the catalytic activities of the enzyme product itself. Establishing the utility of protein fluorescence labeling in cell extracts, in addition to minimal expression systems, (71) advances the investigation of these confounding relationships. It also provides insight into how protein size influences CFE requirements. These in turn enable the optimization of cell extracts for NRPS CFE. Here, a fluorescence-based protein detection method that employs a TC tag and FlAsH fluorophore was adapted for measuring NRPS expression in lysate-based CFE systems. This strategy can validate full-length NRPS translation in the lysate background with higher throughput than current NRPS detection methods. The approach allows the quantification of NRPS expression across several reactions with the use of careful controls that account for caveats of the dye in different reaction environments. Importantly, these measurements enable reaction optimizations that apply to the expression of other NRPSs as well. Overall, the method enables rapid hypothesis generation and testing for broad applications in lysate-based NRPS CFE that could be extendable to other proteins. As shown here with BpsA and its pigment product, this approach can also be complemented with an assay for indirectly measuring improvements to functional protein folding and post-translational modification. (128)

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

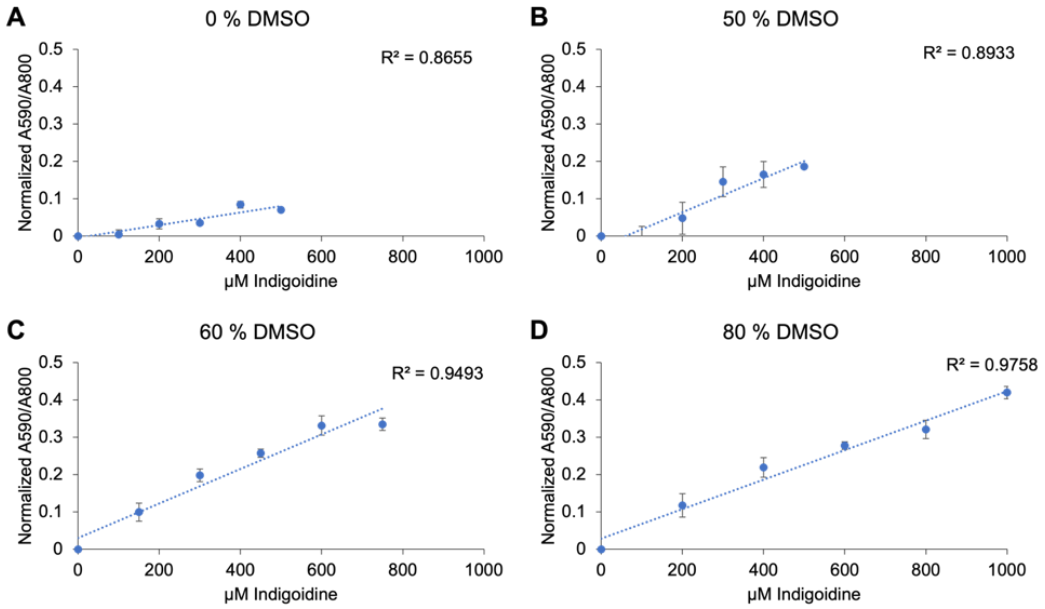
Measuring TC/FlAsH is also a powerful strategy for tailoring lysate-based reaction environments towards high-yield NRPS CFE. For example, throughout this investigation, the TC/FlAsH method identified better temperature, PEP, and magnesium conditions for monomodular NRPS CFE while providing insight into potential bottlenecks of productive amino acid utilization. The results of these experiments lend to further improvements in NRPS CFE systems. For example, the drastic inhibition of NRPS synthesis beyond a threshold concentration of PEP may be remedied using alternative ATP recycling schemes to prevent excessive free phosphate accumulation. (174,189) Factors limiting productive amino acid utilization in NRPS CFE systems can also be further investigated. Apart from

amino acid oxidation, it is also possible that concentrations of other reaction components such as tRNAs or aminoacyl-tRNA synthetases may be insufficient for full-length large protein synthesis even when amino acid levels are high enough. (183) Moreover, extensive efforts to optimize lysates for high-yield production of other NRPS architectures would benefit from TC/FlAsH. While the results reported here are informative for monomodular NRPS CFE, the NRPS family of proteins is known to have high structural diversity. (30) NRPSs with different sizes and numbers of A domains would therefore likely have unique optimal reaction condition sets. The TC/FlAsH method would be useful for defining these reaction parameters through sequential optimization experiments, in a way that offers more throughput than current methods to detect these enzymes. (77,78,93)

Additional applications of the TC/FlAsH approach for NRPS CFE include discovery and prototyping. Despite their natural abundance, difficulty expressing NRPSs in *in vivo* systems has been a barrier to the characterization of these enzymes for natural product discovery. (77) The strategy could thus be usefully applied towards expressing libraries of NRPS sequences incorporating TC tags or prototyping genetic parts for successful NRPS expression in model or non-model lysate backgrounds. Measuring TC/FlAsH could also have value in screening engineered libraries, facilitating the discovery of structurally diverse nonribosomal peptides. (190)

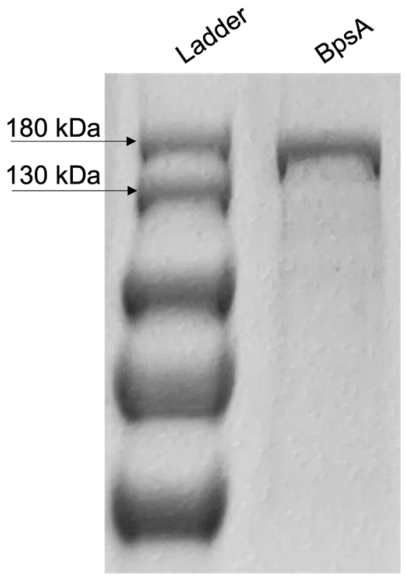
Looking forward, this approach is envisioned to help expedite the study and optimization of lysate-based CFE systems for specific protein synthesis, especially enzymes like NRPSs that possess complex physicochemical properties which may impede cell-free TX/TL. This will consequently advance the development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

Appendix

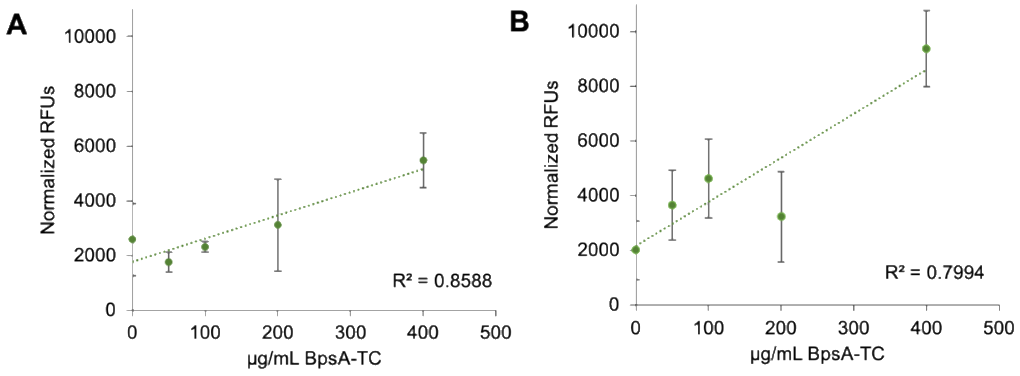


Appendix Figure S3.1. Standard curves of indigoidine from purified indigoidine in BAP1 lysate mock reactions.

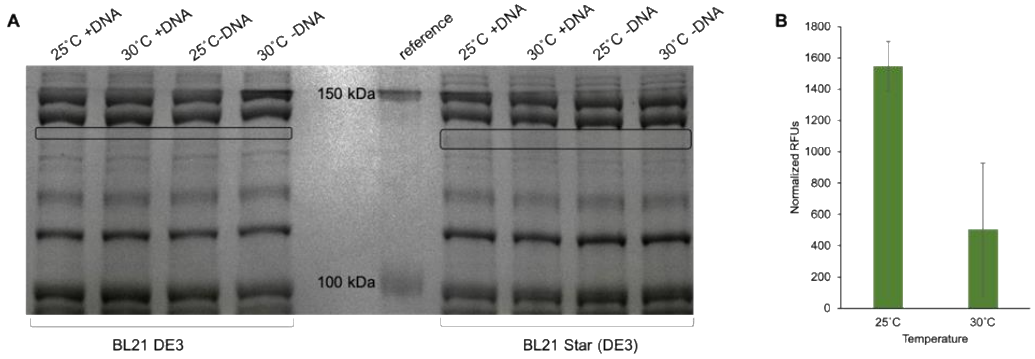
Due to the solubility limits of indigoidine, different concentrations of DMSO were added to the lysate reaction mix to dissolve indigoidine: **A)** 0% DMSO, **B)** 50% DMSO, **C)** 60% DMSO, **D)** 80% DMSO. In all cases, absorbance measurements were taken at 590 nm and divided by signals at 800 nm to deconvolute pigment absorbance from lysate turbidity. (153) Afterwards, A590/A800 signals before DMSO addition were then subtracted from measurements after 70 % DMSO addition to measure soluble pigment. (171) For visualization of trends, values were averaged, compared to the baseline condition (0 µM indigoidine), and plotted with error bars representing the standard error of the mean (n=3).



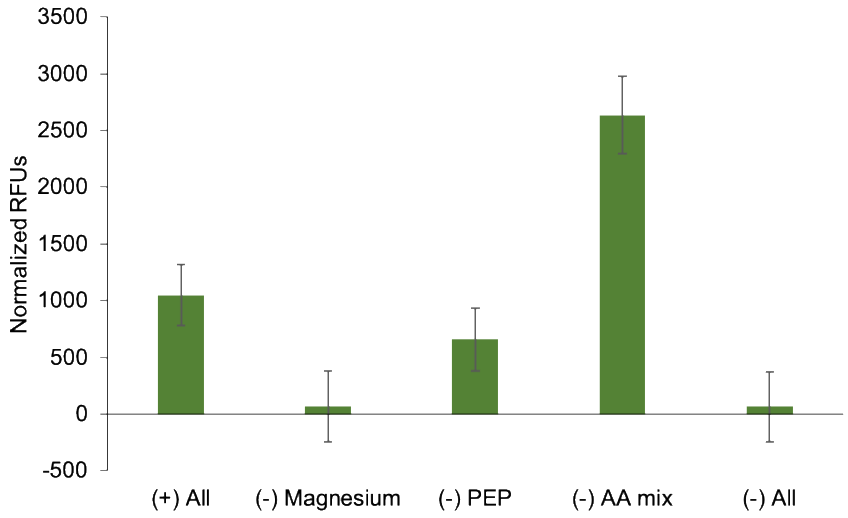
Appendix Figure S3.2. SDS-PAGE gel of purified BpsA-TC expressed heterologously in *E. coli* BAP1.



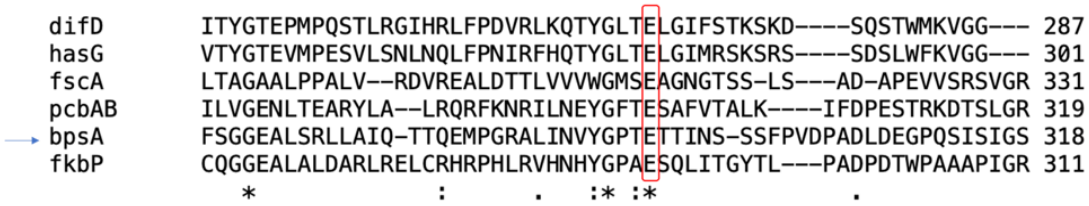
Appendix Figure S3.3 Optimization of FlAsH concentration for TC/FlAsH measurement and validation in *E. coli* lysate-based reactions.
Increasing concentrations of purified BpsA-TC were spiked into BL21 (DE3) lysate-based CFE mock reactions. 0.5 µL FlAsH was added to different final concentrations for each set of reactions to determine an effective FlAsH concentration for measuring a range of TC-tagged protein in lysates: **A)** 7.5 µM, **B)** 15 µM. Averaged normalized values are visualized with error bars representing the standard error of the mean (n=3).



Appendix Figure S3.4. BpsA-TC CFE in different temperatures and E. coli lysates. **A)** SDS-PAGE gel of CFE protein at 25˚C and 30˚C of BL21(DE3) lysate and BL21 Star (DE3) lysate using the CFE base conditions. Lanes represent pooled reactions (n=3). A faint band is most visible compared against the control for BL21 Star (DE3) reactions started with DNA and incubated at 25˚C. Difficulty expressing other NRPSs in BL21(DE3) lysates has been reported and is presumably due to the instability of NRPS mRNA in this system. (77) **B)** Comparison of BpsA-TC synthesized in BL21 Star (DE3) lysate reactions under different incubation temperatures (25˚C vs 30˚C), analyzed with 30 µM FlAsH. These results reinforce higher expression of BpsA-TC in BL21 (Star) DE3 lysate reactions incubated at 25 ˚C, compared to 30 ˚C, which is consistent with reports of higher soluble BpsA expression in cells at lower temperatures. (128) Reactions were run in triplicate and TC/FlAsH measurements were obtained by adding 30 µM FlAsH reagent to each reaction after ~ 20 h incubation. Averaged normalized values are visualized with error bars representing the standard error of the mean (n=3).

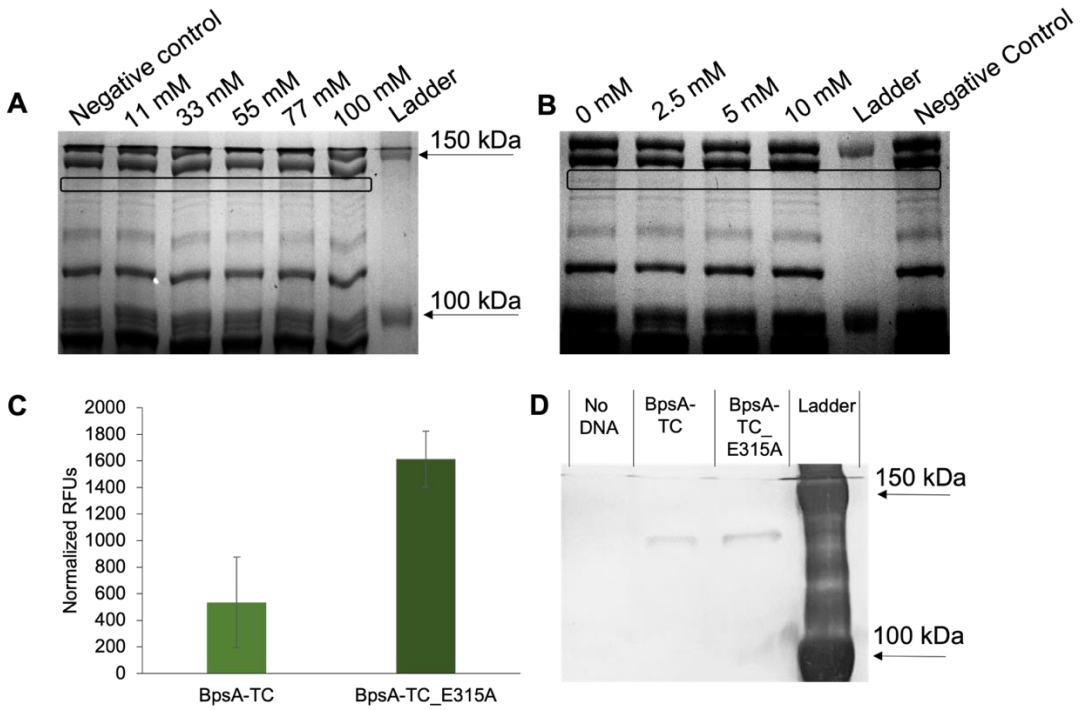


Appendix Figure S3.5. TC/FlAsH measurements in different reaction backgrounds. Magnesium, PEP, the amino acid (AA) mix, or all of these components were removed from mock reaction mixes containing 100 µg/mL purified BpsA-TC. Appropriate controls were prepared, and all samples were distributed in triplicate on a 384-well microplate. The mixes were incubated at room temperature for 1 h. Afterwards, 30 µM FlAsH reagent was added to each replicate and all wells were read with the appropriate filter set after 10 min. Averaged normalized values are visualized with error bars representing the standard error of the mean (n=3).



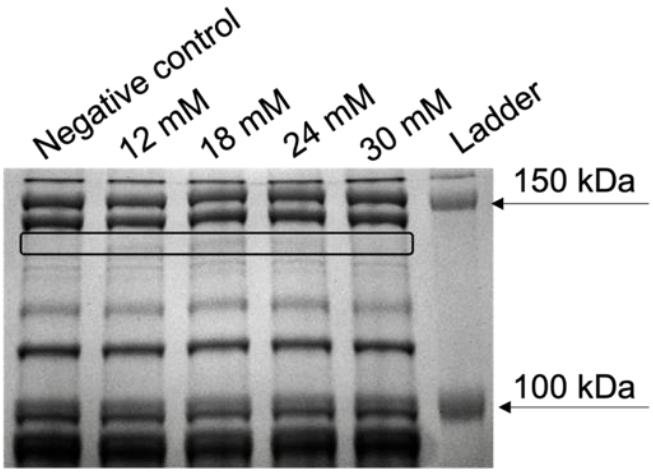
Appendix Figure S3.6. Sequence alignment of BpsA indicating the conserved glutamine involved in the ATP binding motif.

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

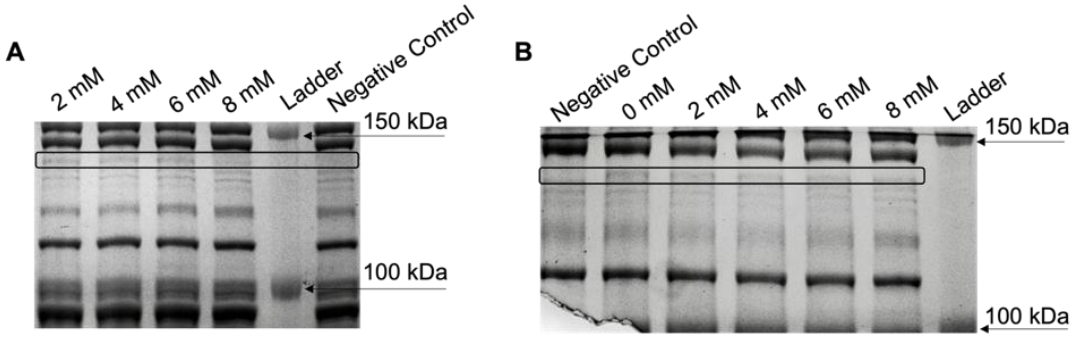


Appendix Figure S3.7. Effects of A domain activity on BpsA's PEP usage and expression.

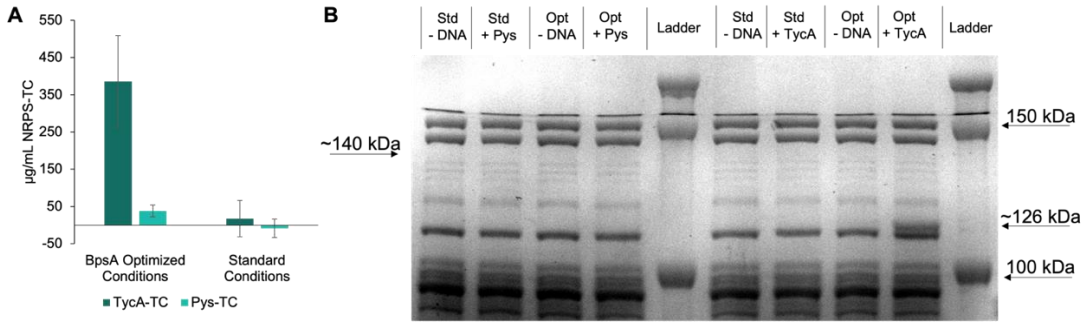
SDS-PAGE gels of pooled reactions (n=3) expressing BpsA-TC prepared with different concentrations of **A)** PEP and **B)** ATP. **C)** Differences in TC/FlAsH signals of BpsA-TC and BpsA-TC_E315A expression in reactions initiated with 77 mM PEP. **D)** Western Blot confirmation of BpsA-TC and BpsA-TC_E315A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



Appendix Figure S3.8. SDS-PAGE validation of magnesium concentration effects on BpsA expression.
SDS-PAGE gel of pooled reactions (n=3) prepared with varying magnesium glutamate concentrations, expressing BpsA-TC. All reactions were prepared with *E. coli* BL21 (Star)DE3 lysates.



Appendix Figure S3.9. SDS-PAGE validation of amino acid concentration effects on BpsA expression.
SDS-PAGE gels of pooled CFE reactions (n=3) expressing BpsA-TC with **A)** varying amino acids concentrations, and **B)** varying concentrations of supplemented L-glutamine. All reactions were prepared with *E. coli* BL21 (Star)DE3 lysates.



Appendix Figure S3.10. Expression of TycA-TC and Pys-TC in lysate reactions prepared with standard conditions or conditions optimized for BpsA.

A) Concentrations of TycA-TC and Pys-TC in optimized or standard conditions, estimated using TC/FlAsH measurements and BpsA-TC/FlAsH standard curves. Separate standard curves were prepared for optimized and base conditions. **B)** SDS-PAGE gel of pooled lysate reactions expressing TycA-TC or PysA-TC under optimized or standard conditions. TycA-TC and Pys-TC expression reactions were all performed in quadruplicate in *E. coli* BL21 (Star)DE3 lysates. "Opt" = reactions run in BpsA optimized conditions (25 ˚C, 77 mM PEP, and 18 mM magnesium); "Std" = reactions run in standard conditions (30 ˚C, 33 mM PEP, and 12 mM magnesium).

Table S3.1. Plasmid and gene sequences used in this study.

Name	Sequence
bpsA E. coli codon optimized part 1	atgacgttgcaagagacttcggtactggagcctaccttacaggggacaacgaccttacccggcttactt gcacagcgtgtcgctgaacacccagaagctattgcggtggcatatcgtgacgataagttaaccttccg cgagttagcttcccgcagcgcggcgttggcagattacttagagcatttaggggttagcgcggatgact gtgtaggcctgtttgtagagccatccatcgacctgatggtgggcgcctggggaattttgaatgcaggag cagcttacttaccccttcgccggagtaccccgaggatcgtcttcgttatatgatcgagaattcggaaac aaaaattattttggcccagcaacgtctggttagtcgcctgcgcgagcttgcacctaaggacgttactatt gtcacattgcgtgaaagcgaagcatttgttcgccctgagggtactgaggcaccagctgcccgtagcgc acgtccagacacccttgcctatgtaatctatacctctggatctacgggaaagccgaagggtgtgatgatt gagcatcgctccattgtaaaccagttaggatggctgcgcgaaacatatgccatcgatcgttccaaggtc atccttcagaaaaccccgatgtcttttgacgcagcgcaatgggagatcctgtcgccggctaacggtgct accgtcgtaatgggcgccccaggagtgtatgcagatccggaagggctgattgaaacaatcgtcaagc acaacgtcacaacattacagtgtgtcccaactctgctgcagggcttaattgatacagagaaattcccgg agtgcgtatcactgcagcaaattttctcgggtggagaggccttaagtcgcctgttagcaattcaaacaac ccaggaaatgccgggacgtgcactgattaatgtatatggccccacggagacaacgatcaatagttcta gctttccagtcgatcccgcggatctggatgagggaccccaatcgatctcaatcggatcgcctgtgcatg ggacgacttaccatatcttggacaaagagactcttaagcctgtgggagtaggcgaaatcggggagtta tacattggaggcattcagctggcccgcggctaccttcatcgtgacgacttaactgccgaacgctttcttg agattgaactggaggaaggtgccgagccggtacgtttatacaagactggagacttaggccaatggaat aatgacgggacggtccagttcgccggtcgtgccgacaatcaagtgaaacttcgtggatatcgtgtgga attggacgagatctctttagctattgaaaatcacgattgggtgcgtaatgcggctgtaatcgtgaagaat gatgggcgcacaggttttcagaatcttatcgcctgcattgaattatccgagaaggaagccgcccttatg gaccagggcaaccacggctcccaccatgcatccaagaaatcgaagttacaagtaaaggctcaactgt ctaacccaggactgcgtgacgacgcagaactggcagctcgtcctgcttttgacctggagggcgctga gcctacgccagaacagcgcgcccgtgtatttgcccgcaagacttaccgcttttacgaaggggagcc gttacgcaggcggatttacttgggcttctgggggctaccgttacggcagggtattcgcgtaaagccgc agatttggcgccagccgaattaggacaaattctgcgttggtttgggcagtacatctcggaagaacgctt gttgccaaagtatgggtatgcatccccgggggctttgtatgctacacagatgtactttgaacttgaaggt gtgggtgggcttaaaccaggctactattactatcaaccggtccgtcatcaactggtgcttatttcagagc gtgaggctactggtaaggctaccgctcagattcactttattggcaaaaaatcgggcatcgagcctgtgt acaagaacaacatcctggaggttttagagatcgaaacgggtcacatggtaggacttttgaacaaatttt gccggcctacggattggacatccacgatcgtgcttacgaaccagctgtgaaggatttgcttgatgtagc tgatgaagactattatcttgggacattcgagttggtgccacatgccggggcccgcgatgaccaagccg aagtgtatgtccaaacgcatggtggtaaagtggccggactgcccgaggggcaatatcgttacgaaaat ggagaacttactcgcttttctgatgacattgtcttaaagaagcatgttattgctatcaaccagtccgtgtac caggctgcatcg
bpsA E. coli codon optimized part 2	tttggaatttcagtatatagtcgcgcagaggaggaatggctgaaatatattacactgggcaaaaaacttc aacatcttatgatgaacggcttgaacttgggattcatgtccagcggttattcgtccaagactggcaaccc cctgccagcttctcgccgtatggacgctgtattaggcgcgaacggggttgactcggccccaatgtatttt tttgtaggaggacgtatctctgatgagcagattggacatgaaggtatgcgtgaagacagcgttcacatg

Table S3.1. Continued

bpsA E. coli codon-optimized part 2 (continued)	cgcgggcctgccgaattaatccgtgacgacttggtctccttcctgccagattatatgatcccgaaccgtg tcgttgtgttcgaccgtttaccgttgagtgctaacggaaaaattgacgtcaaagccttagctgcatcaga ccaagtcaacgcagaattggtggagcgcccgttcgtagcgcctcgtaccgagacagagaaagaaatt gccgccgtatgggagaaggctcttcgccgtgaaaatgcctcggttcaagacgacttcttcgaaagtgg aggtaattcgctgattgcagtaggattagtccgcgaactgaatgcacgtttaggcgttagccttcccttac agtctgtacttgagagccctaccattgagaagttggctcgtcgcttggaacgcgaagtcgcgcaggaa tcgtcccgctttgtgcgtcttcacgctgaaaccggcaaagcccgtccggtcatctgctggcctggtctg ggaggctatccaatgaatttgcgctcacttgcaggagagatcggcctgggacgctcgttctatggggtt caaagctatggcatcaatgaggggagactccctacgaaaccattactgagatggcgaagaaagaca tcgaggccttgaaggagatccagcccgcagggccatatactctgtgggggtattctttcggcgctcgtg tggcgtttgagactgcatatcagctggaacaggcgggcgagaaagtggacaacttattcttaatcgccc cgggtagtccaaaagtccgtgcggagaatggcaaagtctggggccgcgaggccagtttcgccaatc gcgggtatacaacgatccttttctccgtctttactgggaccatcagcggcccagatcttgatcgctgctta gagacggtaactgacgaggcatccttcgccgaattcattagtgaactgaaaggcattgacgtggacttg gcgcgccgtattattagtgtcgttggccaaacatatgaattcgaatactcgttccatgagttggcggaac gtactttacaggcacctattagtattttcaaggctgtgggtgacgactattcgttccttgaaaattcgagcg gctatagcgctgagccaccgaccgtaattgacttggacgccgaccactatagcttgttgcgtgaagac atcggagaactggtcaagcatatccgctatttacttggtgaatgttgcccgggttgctgtcatcaccatca ccatcactaa
pys E. coli codon optimized part 1	atgcgcgatgtccacccatgcccactgagtactgcgcaacaagaagtgtgcttggccattagtcaatct cattcaaatctgaattaccatctttgtgatgtaattgaattgcgcggcgaacttcacttgccttggcttgaag cggccattcgtgcccagtttttacaaacggacacccttcgcgcaacattcgacatcgacccgactacgg gaacatacgcccagcatatccagcctgccgatgcgttaccggcgaaggcatttgagtcacttgacgta agcgaccaggcagaccccgcgcaggccagtaacgagcttttggagcatctgttggatcaagacatgg atttacaacacgggccgttaatccgctatgtgctgattcgtctggcaccgcaacattatcgcatgattgaa ctggcttctcatttagtagtcgacggcttcggccacggaatccttttgggaatattactgctcactataac gctttgtcacgcggtgagacggtagaagcacttgaattagccccgttatctagcgtgttcgatgcccaa gaagagtatcgccattcggatatgcataagaaggatcgtacgtattggcgccagtattgtctgaaaatg cccgaaccaacacaattggtccctggagatgctccgttgattaaactgaatcgtctgcgcaaagtctttg gaggcgccacacttcttcagcttcgtgccgcggcgtccgagcatcaacttcgtttatcctctatcctgttg gctttatgtgccacatatttgcagcgcatgaccgggcagcacgagttggccttgggcatgccagttgcg gcacgtcagttaaaggctttgcgcaatgtaccctctatggtagcaaatatcttgccactgcacttgtcattt acaccggagtcaaccgtcttaagcgtcgcggccaatcttcaacgccagttacgccatcacttgctgcat caatcttaccgttcggagtcgatgatccgcgacttacatgcggaacgcgggaataagccgttatttaata ccttgttgaacatcgttgcgtatgaccaagggccgggatttgcgggctgcgatactacgatccaaaacg tcgccaacgggccagccgaccaccttggcattgacatcttcgaccgtcatgacgacggtcgtttggag atcgggttcaatgccaatgctgatctgtattcggctgaggcattggagctgcactaccaacgtttgactg cgttattcgagcgcttcgctaacgctccccaaacgctggcagcagactataacctgttcttaccagacg agcaacaacgccactacgcacttcccccccatccagaaaagcttcctgtttttgcggaggcgtttgcga acgcggttcatgactacgctgagcgcccagcactttcccaaggagcccagaccctgagttaccgtcgt

Table S3.1. Continued.

T*pys E. coli* codon optimized part 2	ttggacgatgacgctacacgcctggcagcccacctgcgcgaacgtggggtgcgcgccggagattgc gtcgtagtcatgttcagtcgctccgtcgagtgggcggtagcggccgtagctttgtttaagctgggagcgt gttacgtaccagtggatccagatctgcccgaagcccgcattgaacatatcttctctgacgccgacccgg cagtagtcattgtggcaccaggatcacagcttaaggtggaggtagccgcagataaactgctgcgtctttc cgcagagactttagctcagttgccagctgtcacgcaacctctgtcagcattcgatgctggcttgcctggct atcttatttatactagcggatctacagggaagcctaagggcgtggaggtgacgcagcgcaacttagtgc ccatcgctcgtactgctatcaatgccgcccagctgcaaccaggcgcgcgtgtgcttcagttcatcgcag ctggatttgacatgtctgttttagagattatgatgacgcttttagcgggggccgaattagtaattacagaca aagtatcgagtgctccagggaaagcgcttgctacactggttaagcgtgagggcatcaatttgcttgtaat gaccccgtccctgttggcctgtcaccaaactgaggactttccccaggatacggtcttaatgcttggaggg gagccttgtacgcccgccctgctggcgcgttttgcccactgtcgtttgttgaacgtgtatggaccaacag aaaccagtttcgcaacaagtattaacgcctgctatggagcgggggacttatctatcggtccggccacgg ctaacactcgcttgtatgtagtcgatgggcaacagcgcttacttcctccgggagcctgggggggatttgttt atcgggggtcccggtgtagcccgtggctaccgtaaccgccccgacttaactgctaagggttttgtagca gatttgttagactctgcatcaacgatgtaccgtgcaggggaccgtgtgttcttcgaccatgccggacgca ttcattacctgggtcgtcaagataatcaaattaagttacgcggtttgcgcatcgaattagacgagattaaga atgtcttgttaggatgtgcaggcgtagatgacgccacagttattttacgtgaattggggcacggaccagc gattgtgggttacgtggcgtctaaggattccagtcttgatggtcaacatcttaagcaggcgctgggccgt catttaccccaacatatggtgccgtccgtaatcatgtgtgttagcgagttccctctgacccccaatggtaaa ttagctgttgatcgtttgccagctccggcactgtatgatgccactgaactggctccagctcaaaacgctga ggaggcggccatgtgcaagttgtttgcggaggccttagattgcaacgaagttttcgcgaaccagaatttt ttgaccttggcggccatagtttactggggttgcagcttgtgagtcgtattaaggagcagttcgggattgca cttggaattgccgattttcttgctgcgccaacaccgcgccaattagcgcagcgcttacagacaaactcag gcgactctgacccgtttgatgcagtcttaaccctgcgcagtgaaggttctcgcccacccctgttttgcatc catccaggggggggattgcgtggccctatgccggattattaccattcttgccggaagaccaaccaatg tacgcccttcaatcgccgattctgcgtgatccgactcgtgttatcgggagcttagatgaactggccgccg agtatttacaacgcattgtagacttgcatcctgaaggtccctaccaacttgcaggctggtcagtgggagg caaccttgccttacgtatcggtgctatgcttcaagcgcgtggccgtgaggtatctttttttatgcatgtttgact catatccattgcagggaggccccgcttcgttgaaacttgatgatgcgatgattatctcacgcatgacccgt gctattgtcggaactcctcgtgcgggattaaagggcctgaagtccgcaatggaagaagtcttgggaagc cgtcaaattggcgacgaattcttgacacgccttgtagacgattcaaaattgatgttagagttgttagggcg cacgcagtatgaggtatttaatggcgatctgctgttcattcgcgcgacaacagacattcttcgccaagatg agcagcagccaggtctgtgggcaccgtacatttctggtgagttaatccagcacgacgtcgaggctcctc acgaatgcttgcttcaacgccagtacttagaacaatttgggaaggcgttcgtagaggctttacttaaacgt caagcctcccccgcttcggaaacccaaaagctgtcggaagtctgttgcccgggttgctgttggagccac ccgcagttcgaaaaataa
tycA E. coli codon optimized part 1	atgcttgcaaaccaagctaatcttattgaccttgaatgggagaaagaacaagttgcgctttttaacgacac caaagcggagtaccctcaggacaagagcatccaccaattattcgaggaacgtgccgaagcagtcccg catcgcgtggcaattgtgtatgaaaatgagcgcttaacctaccaggaattaaataaaaaagccaaccaat tagcgcgtgccctgattgaaaaaggcatcacaacagattctatcgtgggagtcatgatggaaaagtctat

Table S3.1. Continued.

tycA E. coli codon optimized part 1 (continued)	tgaaaatgtcacagccattttagctattttaaaggcaggtggagcgtatgttcccatcgacattgagtaccc gcgtgaccgtatccaatacatcttgcaggactcccagactaaaatcgtaattacacaaaagaatgttagtca attagtctacgatgtcgggtatcgtgggcctgttgtcgtattggaagaagaacagctggactctcgtcctaa ctcgaacttgcatttgccgtcgaaaccgactgatatggcctacgtaatttatacctcaggcacaacaggca agccaaagggaactatgttggaacacaaaggtattgcaaacttgcagtctttttttcagaacacgtttgacgt aacgcccgccgaccgtattggtcagttcgcgtccatgagtttcgatgcaagtgtctgggagatgtttatgg ctcttttgaccggagccagcttatatattttgtcgaaacagacgattcatgacttcatcagcttggaaaactac cttaatgagaatgagcttaccattattactcttcccccgacataccttacacaccttaatcctgaccatattact actcttcgtatcatgattacggccggatctgcagcaaattttccgttagtaaatcgttggaaaaataaagtcc gttatgtaaacgcctatgggcctacggaaacaagcatttgtgcgactatttgggaggcgcctcctaacgaa cttttaaatcagacaatccctatcggaaaacccattcagaatacatatgcatatattgttaacgaagaactgc aacttcagcctattggctcggaaggggaattgtgcatcgggggagtagggctggcccgcgggtattgga accgtccggatttaaccgcggagaagtttgttgataatcctttcgttccaggggaaaaaatgtatcgtacag gggacctggccaaatggcttccgtctggagagatcgaattccttggccgtatcgaccaccaggtcaagat ccgtgggcaccgcatcgagttgggcgaaattgagtctgtattgcttaaacatgaacaaattaaggaagctg tcgtaattgcgcgcgaagaccaacatgcgcaaccatatttgtgcgcatacttcatttcgcagaaagaagtg acgcccgctcaaattcgtgaatacgccgctcagaagctgccagcatacatgttgccgtcttatttcgttaaa cttgacaagatgccgctgacacctaatgataagattgaccgcaaggccctg
tycA E. coli codon optimized part 2	cccgaacctgacgtaacgtcggaaactaatccagcgtatgagcctcctcgcaatcagaccgaaaagatt ctggctacagtatggcaggaagttcttggaatcgaaaagattgggattctggacaacttctattcgttaggt ggtgatagcattcaagcgattcaagtggcagcgcgtctgcacgcctatcagcgcaagttagacacgaag gatttattaaactatccgaccatttcccaagtcgcattatacgtgaagtccacgacccgcagctccgaacag ggaattatcgaaggcccagtagcactgacacccatccagcactggtttttcgagaaaaacttcacaaacat gtcgcattggaaccaaagctatgttctgtatcgtgaacaggggttcgatccagaggccattcgcaaagctc tggataagattttaagtcaccacgacgcgttgcgcatgatctaccagcgtgaaaatgggtctatcgtacag cagaatcgcggattgggtagtgaactgtatcaattcttgcaatatgacctgacatcgcactgggacgtaca gaaggctatcgaagaagaaaccaaacgcttgcatagttcaatgaatctgcaagaagggcaattagtaaaa gcggccctttccacacgcttcaaggcgatcacttgttccttgcaatccaccatctggttatggacggtattt cttggcgcattctgttcgaagaccttgcaacggcatacagtcaggtagttgcggggaaggagatcgttctg cccgagaaaactgattcatttaaggactggacagtgtggttaaacgattacgctaatagcgacgctctgat gcgtgaaatcccctactgggagaacttggaatcaggagctcgtaatgcctcacttcccaaagactatgag acagtaggatgtaagcaaaaaagtacccgcaatatccaggttggtcttctggcagctgaaaccgaacaat tactgaaacaggcaaatagcgcttaccaaacggaaattaatgatctgctgttagccgccctgggtcttgca gtggctgagtggggacgtcttgaccaaattgttatcaacttggaagggcacggtcgtgaagacgttattga gcacgcaaatgtgactcgtactattggctggtttacatctcagtatccggtccttctggacgtctctcaaatg aaccacttaccaaaccacattaagtgtacaaaggagaatcttcgcaaaatccctaacaagggtatcggcta cgaaatcttaaagtatatgacggaacccgaacgtcgcggctctttgtcattctctctgcaacctgaggttac attcaattatttaggtcagtttgattcagggttgaattcagaactgtttactcgtagcccatactccagcggga ataccttggagccgatggtaagaataatcttagtccggattctgagatttacactgccctgaacattaccg gacgcattgaaggtggagagcttttgatcacatttacgtattccgaggagcagtttcgtgaggagtccatcc

Table S3.1. Continued.

tycA E. coli codon optimized part 2 (continued)	atcaattatcggaaaattaccagaaacacctgcgctcgatcattacacactgcgtacaaaaaaaagaagta gagcgtacaccctcggatttctccctgaagggccttcagatggaggaaatggatgatatctttgaggtcttg gctaacaccttaaactgttgcccgggttgctgttggagccacccgcagttcgaaaaataa
pET28a(+) sequence with TC tag (continued)	gatgtaactgaatgaaatggtgaaggacgggtccagtaggctgcttcggcagcctacttgttgagtagag tgtgagctccgtaactagttacatcctagcataacccctttggggcctctaaacgggtcttgaggggttttttg cggatccgaactcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaa gctgagttggctgctgccaccgctgagcaataactagcataacccctttggggcctctaaacgggtcttga ggggttttttgctgaaaggaggaactatatccggattggcgaatgggacgcgccctgtagcggcgcatta agcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcc tttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttt agggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgg gccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttcc aaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctatt ggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggt ggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctc atgaattaattcttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaatac catatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaaga tcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataa ggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagtttatgcatttctt tccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattca ttcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcga atgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatac ctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgct tgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaa cgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcac ctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcgg cctagagcaagacgtttcccgttgaatatggctcataacacccttgtattactgtttatgtaagcagacagtt ttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatca aaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagc ggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagat accaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacata cctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactc aagacgatagttaccggataaggcgcagcggtcgggctgaacgggggggttcgtgcacacagcccagc ttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttccc gaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgaggg agcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgat ttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctg gccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctt

Table S3.1. Continued.

pET28a(+) sequence with TC tag (continued)	tgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcgga agagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatatggtgcactctcagta caatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgc cccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaa gctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctg cggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgag tttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactg atgcctccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgat acgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcg gcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacaggg tagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttccagactt tacgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttc acgttcgctcgcgtatcggtgattcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaa cgacaggagcacgatcatgcgcacccgtggggccgccatgccggcgataatggcctgcttctcgccgaaa cgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgaca ggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctg tcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccgga aggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaactta cattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggc caacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgagacgggcaa cagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcag gcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactac cgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgt tggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggc actccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagac gcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccag atgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacat caagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagt taatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcg ttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgac ggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgc cacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggct ggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttac tggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccat tcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttga ggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtcccccggcca cggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttcccc atcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtc cggcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataa caattcccctctagaaataattttgtttaactttaagaaggagatatacc

Table S3.2. Primers used in this study; highlighted nucleotides are for the A domain point mutation: E315A.

Primer name	Sequence
E. coli BpsA part 1 Forward	CACACCAGGTCTCACAATGACGTTGCAAGAGACTTCGG
E. coli BpsA part 1 Reverse	CACACCAGGTCTCAAACGATGCAGCCTGGTACACGG
E. coli BpsA part 2 Forward	CACACCAGGTCTCAAACGATGCAGCCTGGTACACGG
E. coli BpsA part 2 Reverse	CACACCAGGTCTCATTATTCACCAAGTAAATAGCGGATATGC
pET28a with *E. coli* BpsA Forward	CACACCAGGTCTCAATAACTGCCAGCGCATCAACAATATTTTCACC
pET28a with *E. coli* BpsA Reverse	CACACCAGGTCTCAATAACTGCCAGCGCATCAACAATATTTTCACC
BpsA-TC_E315A part 1 Forward	AGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGT
BpsA-TC_E315A part 1 Reverse	ATCGTTGTCGCCGTGGGGCCATATACATTAATCAGTGCA
BpsA-TC_E315A part 2 Forward	GTATATGGCCCCACGGCGACAACGATCAATAGTTCTAGCTTTCCAGT
BpsA-TC_E315A part 2 Reverse	CGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACG
E. coli Pys part 1 Forward	CACACCAGGTCTCAGCATGCGCGATGTCCACCCATGC
E. coli Pys part 1 Reverse	CACACCAGGTCTCAACTAAGTTGCGCTGCGTCACCTCC
E. coli Pys part 2 Forward	CACACCAGGTCTCATAGTGCCCATCGCTCGTACTGC
E. coli Pys part 2 Reverse	CACACCAGGTCTCATTTTATTTTTCGAACTGCGGGTGGCTCC
pET28a with *E. coli* Pys Forward	CACACCAGGTCTCAAAAAGCTTGCGGCCGCACTCG
pET28a with *E. coli* Pys Reverse	CACACCAGGTCTCAAAAAGCTTGCGGCCGCACTCG
E. coli TycA part 1 Forward	CACACCAGGTCTCACAGCATGCTTGCAAACCAAGCTAATCTTATTGACCT
E. coli TycA part 1 Reverse	CACACCAGGTCTCAGGCAGGGCCTTGCGGTCAATCT

Table S3.2 Continued

E. coli TycA part 2 Forward	CACACCAGGTCTCAGGCAGGGCCTTGCGGTCAATCT
E. coli TycA part 2 Reverse	CACACCAGGTCTCATTTTATTTTTCGAACTGCGGGTGGC
pET28a with *E. coli* TycA Forward	CACACCAGGTCTCAAAAAGCTTGCGGCCGCACTCG
pET28a with *E. coli* TycA Reverse	CACACCAGGTCTCAGCTGCTGCCCATGGTATATCTCC

CHAPTER 4

PROFILING EXPRESSION STRATEGIES FOR A TYPE III POLYKETIDE SYNTHASE IN A LYSATE-BASED, CELL- FREE SYSTEM

A version of this chapter has been submitted for publication and has been submitted as a preprint to Biorvix by Tien T Sword, Jaime Lorenzo N. Dinglasan, Ghaeath S. K. Abbas, J. William Barker, Madeline E. Spradley, Elijah R. Greene, Damian S. Gooden, Scott J. Emrich, Michael A. Gilchrist, Mitchel J. Doktycz, and Constance B. Bailey:

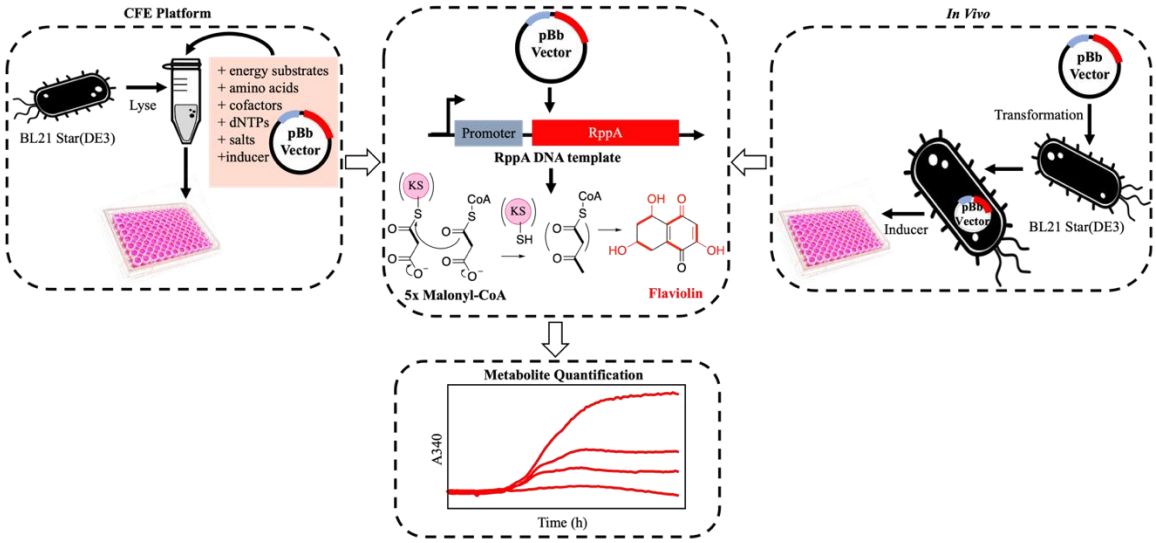
Sword TT*, Dinglasan JLN*, Abbas GSK, William Barker J, Spradley ME, Greene ER, et al. Profiling Expression Strategies for a Type III Polyketide Synthase in a Lysate-Based, Cell-free System. BioRxiv. 2023 Dec 1; Under Revision in *Scientific Reports*

*Authors contributed equally

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

4.1. Abstract

Some of the most metabolically diverse species of bacteria (e.g., Actinobacteria) have higher GC content in their DNA, differ substantially in codon usage, and have distinct protein folding environments compared to tractable expression hosts like *Escherichia coli*. Consequentially, expressing biosynthetic gene clusters (BGCs) from these bacteria in *E. coli* frequently results in a myriad of unpredictable issues with protein expression and folding, delaying the biochemical characterization of new natural products. Current strategies to achieve soluble, active expression of these enzymes in tractable hosts, such as BGC refactoring, can be a lengthy trial-and-error process. Cell-free expression (CFE) has emerged as a valuable expression platform for enzymes that are challenging to synthesize *in vivo*, and as a testbed for rapid prototyping that can improve cellular expression. Here, we use a type III polyketide synthase from *Streptomyces griseus*, RppA, which catalyzes the formation of the red pigment flaviolin, as a reporter to investigate BGC refactoring techniques. A library of constructs with different combinations of promoters and *rppA* coding sequences was created to investigate the synergistic tune effect between promoter and codon usage. Subsequently, we assess the utility of cell-free systems for prototyping these refactoring tactics prior to their implementation in cells. Overall, codon harmonization improves natural product synthesis more than traditional codon optimization across cell-free and cellular environments. More importantly, the choice of coding sequences and promoters impact protein expression synergistically, which should be considered for future efforts to use CFE for high-yield protein expression. Additionally, promoter stratergy of RppA does not show the same trend with sfGFP, indicating that



Scheme 4.2. Production of Flaviolin *in vitro* and *in vivo*.

different proteins have various refactoring strategies. *In vivo* experiments suggest that there is a strong correlation, but not complete alignment between expressing in cell free and live cell. Refactoring promoters and/or coding sequences via CFE can be a valuable strategy to rapidly screen for catalytically functional production of enzymes from BCGs, which advances CFE as a tool for natural product research.

4.2. Introduction

Microbial secondary metabolism generates a vast number of complex secondary metabolites, or natural products, with varied chemistry(191). Members of the order Actinomycetales, and especially those from the genus *Streptomyces*, have been shown to be a valuable lineage in terms of their capacity for secondary metabolism(58,97,104,192). Indeed, *Streptomyces* genomes are known for harboring a large number of biosynthetic gene clusters (BGCs) that are attractive for the discovery of novel enzymes and their chemical products. Unfortunately, these positive aspects of *Streptomyces spp.* are offset by their slow doubling time, mycelial clumping, thick cell walls, high GC content (~70-75%), and relatively cumbersome genetics. Consequently, biochemists who study *Streptomyces spp.*-derived natural products struggle to generate enough protein for biochemical or structural characterization in a short amount of time. As a result, most efforts for expressing proteins of interest as soluble, functional constructs are done in tractable hosts like *E. coli*, rather than *Streptomyces spp*. Because of the deep evolutionary divergence between Actinobacteria and Proteobacteria such as *E. coli*, the differences in their metabolic backgrounds, dissimilar codon usage and genome attributes, as well as protein folding environments, expression of heterologous genes from *Streptomyces spp.* and, in turn, product synthesis often fails(101). Where BGC expression in *E. coli* is possible, a myriad of parameters usually require optimization. Often, using the native coding sequence amplified from genome DNA with common expression systems (e.g., the pET expression system)(193),(194,195) results in poor expression or inclusion body formation(94). Additionally, product toxicity can impede the use of *E. coli* strains without engineering host tolerance. Finding suitable recombinant expression systems, therefore, involves screening of refactoring choices that include choice of regulatory elements like promoters, and coding strategy. This process can be time-consuming and laborious, a significant bottleneck to researcher workflows.

Cell-free expression (CFE) platforms employ either crude cell lysates (or extracts) or an *in vitro* transcription and translation (TX-TL) PURE system and can bypass limitations in secondary metabolite production observed *in vivo* expression systems(67,196). The TX-TL PURE system or PURExpress employs the minimal number of recombinant elements required for transcription and translation while approaches using crude cell lysates are derived from intact living cells. Harnessing the TX-TL machinery preserved in lysates allows protein expression in the absence of other normal cellular functions, a feature that can be leveraged to manufacture enzymes that are difficult to

synthesize in microbial hosts. Lysates also retain metabolic pathways that can be engineered to accumulate precursor molecules for heterologous biosynthetic enzymes(177,178,197). CFE systems are thus often used as an alternative approach for synthesizing BGCs and their product metabolites, especially when cytotoxicity represents a limiting factor to heterologous *in vivo* production(61,74,198). Additionally, CFE systems can be used to optimize the cell-based expression of soluble BGCs when leveraged as testbeds for genetic refactorization. Prototyping different genetic constructs in these platforms is relatively rapid as it bypasses time limitations associated with culturing and genetically manipulating live cells. To these ends, investigating refactoring strategies in a cell-free environment can benefit the development of cell-free and cell-based BGC expression platforms.

While optimizations of protein expression via refactoring have usually focused on robustly expressed reporters (e.g., sfGFP)(89,199–203), we sought to evaluate common refactoring parameters using a reporter that was relevant to the enzymatic activity of genes involved in secondary metabolite formation(193). To further explore the ability to refactor for functional catalytic activity, we chose a model protein, RppA (40.1 kDa), that generates flaviolin, which is a red pigment that has limited catalytically functional expression in *E. coli* even with current optimization(204–206). RppA is a type III polyketide synthase from *Streptomyces griseus*, that utilizes acyl CoAs as substrates directly, thus it does not require phosphopantetheinyl transferase to be active. Using flaviolin production as a reporter for catalytically functional RppA expression, we varied parameters that are relevant to improving expression. This included the use of three different, commonly used inducible promoters: the workhorse T7/*lac* system, the pBAD arabinose promoter(207), and the pTet anhydrotetracycline promoter(208). In addition to varying promoters, we also evaluated the impact of four different methods for designing synonymous coding sequences for generating higher levels of heterologous gene expression by using codons that better reflect the host's (*E. coli*'s) codon preferences. Notably, while used extensively *in vivo*, inducible promoters beyond the T7 system have not been extensively investigated in CFE. We also demonstrate strong positive correlations between cell-free and cell-based expression and discuss the feasibility of this approach for refactoring challenging proteins *in vitro* prior to *in vivo* production. Taken together, this work demonstrates a coordinated strategy to apply a lysate-based cell-free environment for profiling genetic constructs that promote catalytically active enzyme formation. The results are applicable to both cell-free and cell-based systems and can be used to generate biosynthetic proteins for characterizing elements of engineered biosynthetic pathways.

4.3. Methods

4.3.1. Strains and plasmids

E. coli BL21 Star(DE3) was purchased from New England Biosciences (Ipswich, Massachusetts, USA). *Streptomyces griseus* was purchased from Carolina (cat# 155705). The *rppA* gene from *S. griseus* was codon optimized using Integrated DNA Technology's (IDT) codon optimization tool (https://HR.idtdna.com). *rppA* codon optimized and *rppA* codon harmonized with a C-terminus strep tag, TGGAGCCATCCGCAGTTCGAAAAA, were ordered from IDT (**Table 6**). BioBrick plasmids were obtained from Addgene (https://HR.addgene.org): pBbE2k (Plasmid #35324), pBbE7k (Plasmid #35315), pBbE8k (Plasmid #35270), and pJL1-sfGFP (Plasmid #102634) (**Table 7**). All of the constructs were cloned via Gibson assembly (New England Biolabs, part #E2611S). Primers used in this study were designed with the J5 algorithm and are listed in **Table 8**.(126)

4.3.2. *In vivo* Flaviolin measurement

E. coli BL21 Star(DE3) was used as a host strain for *in vivo* expression of RppA. Cultures were grown in 2xYPTG media (10 g/L yeast extract, 7 g/L potassium phosphate dibasic, 3 g/L potassium phosphate monobasic, 5 g/L NaCl, 16 g/L tryptone, and 18 g/L glucose) supplemented with kanamycin at 50 µg/mL. Overnight seed cultures (2 mL) were grown from a fresh single colony at 37°C, shaking at 210 rpm. In a 96-well plate (Greiner), all constructs started growing with the initial OD_{600}= 0.005. 5000 µM L-arabinose, 100 nM anhydrotetracycline, and 500 µM IPTG were induced after 110 min, 180 min, and 210 min, respectively. The plate was covered with an adhesive plate seal (Thermo Scientific) and loaded measured on a VARIOSKAN LUX (Thermo Scientific) plate reader. Readings at A_{340}/A_{600} were taken every 10 min for 20 hr.

4.3.3. *In vivo* Flaviolin production and purification

A 25 mL seed culture of BL21 Star(DE3) harboring the optimized *rppA* driven by pT7 plasmid was grown overnight (37 °C, 2010 rev/min) in LB medium supplemented with 50 µg/mL kanamycin. After ~20 hours, 10 mL of seed culture was used to inoculate 1 L media in a Fernbach flask (VWR 29171-854). Cells were incubated at 37 °C shaking at 210 rev/min. At an OD_{600} of ~0.8, the culture was induced with 0.5 mM IPTG and grown at 16 °C for 20 hr. The culture was then centrifuged at 5000 x g for 30 mins. The pink supernatant was adjusted to pH=2 with 3M HCl and incubated at 4˚C overnight to precipitate flaviolin. Pigments were recovered by centrifugation at 5000 x g for 30 mins, and the precipitate was washed with DI water. The pellet was then dried at 50˚C in an oven overnight. The dried pellets were washed with 6M HCl at 100˚C to remove proteins and carbohydrates, then centrifuged at 5000 x g for 10 mins. The precipitate was washed with ethanol and chloroform and then dried at 50˚C overnight. Purity was verified via Direct

Analysis in Real Time Mass Spectrometry - DART-MS (*m/z* = 206), and ^{1}H NMR: (400 mHz, acetone D_6) δ 12.53 ppm (s, OH), 2.29 ppm (m, CH), 6.11 ppm (s, CH), 7.08 ppm (d, CH) and 6.61 ppm (d, CH) which is in agreement with the previous literature report. (250),(251)

4.3.4. Cell-free extract preparation

The same extract preparation procedure was used for all strains. A seed culture was prepared with 30 mL 2xYPTG media (10 g/L yeast extract, 7 g/L potassium phosphate dibasic, 3 g/L potassium phosphate monobasic, 5 g/L NaCl, 16 g/L tryptone, and 18 g/L glucose) inoculated with a fresh colony and incubated overnight at 37˚C, 220 rpm. 1 L of 2xYPTG media in a 2.5 L Tunair flask was then inoculated with the overnight culture and grown at 37˚C, 220 rpm. Cell growth was monitored by NanoDrop (Thermo Scientific). Cells were harvested at OD_{600} ~ 2.8-3.2 by centrifugation (5000 x g, 15 min, 10˚C), then washed three times using S30 buffer (10 mM Tris-acetate, 14 mM magnesium acetate, 60 mM potassium acetate, and 10 mM DTT). All wash steps were performed at 4˚C. Cell pellets were then weighed, flash frozen, and then stored at -80˚C. For extract preparation, the cell pellets were then thawed on ice and resuspended in 0.8 mL of S30 buffer per g of cell pellet of the pellet by vortexing with short bursts (vortex 15 s, rest 30 s, repeat). 1.4 mL aliquots were sonicated on ice in 2 mL microcentrifuge tubes using an OMNI Sonic Rupto 400 (45 s on, 59 s off for three cycles, 50% amplitude set). 4.5 µL of 1 M DTT was added into each tube immediately after sonication. All samples were centrifuged at 12000 x g for 10 minutes at 4˚C. The supernatant was collected without disturbing the pellet and centrifuged again to remove the remaining debris. The resulting supernatants were aliquoted into fresh centrifuge tubes, flash-frozen, and stored at -80˚C.

4.3.5. CFE reaction preparation

The cell-free reaction comprised 1.2 mM ATP, 0.85 mM GTP, 0.85 mM UTP and 0.85 mM CTP; 34.5 µg/mL folinic acid; 0.4 mM nicotinamide adenine dinucleotide (NAD), 0.27 mM coenzyme A (CoA), 4 mM oxalic acid, 1 mM, 1.5 mM spermidine, 57.33 mM HEPES buffer, 10 mM magnesium glutamate, 10 mM ammonium glutamate, 130 mM potassium glutamate, 2 mM each of the 20 amino acids, 33 mM phosphoenolpyruvate (PEP), 27 ng/µL DNA template and incubated at 30˚C.(241) Reactions were set up in 10 µL volumes unless otherwise stated. The type of inducer used and changes to any of these conditions are described in the text. All reactions were incubated in a 96 PCR well plate (VWR #47744-116). Surrounding wells were filled with 1x phosphate-buffered saline (PBS) to control the humidity and prevent evaporation. Plates were covered with an adhesive plate seal (Thermo Scientific), before putting it in the plate reader. Flaviolin synthesis was monitored by reading reaction absorbance at 340 nm at varying timeframes and intervals, as described in the text.

4.3.6. Flaviolin quantitation in lysates with absorbance measurements

To generate standard curves from pigment absorbance measurements, increasing concentrations of the purified pigment dissolved in DMSO were spiked into BL21 Star(DE3) lysate mock reactions (i.e., reactions without DNA). Absorbance measurements were made in a 96 PCR well plate, without a lid, loaded into a VARIOSKAN LUX (Thermo Scientific) plate reader. The read protocol was set to shake the plate at high speed for 2 s then measure absorbance in selected wells at 340 nm. The resulting values were then normalized to the 0 µM pigment condition.

To measure the absorbance of flaviolin produced by cell-free expressed *rppA*, base reaction mixes with BL21 Star(DE3) lysate and RppA-expressing plasmid DNA were performed. Modifications to the reactions for validating pigment production are described in the text. All reactions were laid out on a 96-PCR well plate and measured every 10s for 20 HR at 30˚C. A_{340} measurements were taken and normalized as described above.

4.3.7. Quantification of active sfGFP

Fluorescence measurements of reactions expressing sfGFP were taken using top optics on a VARIOSKAN LUX (Thermo Scientific). Excitation and emission filters were set to 485 nm and 538 nm, respectively.

4.4. Results and Discussion

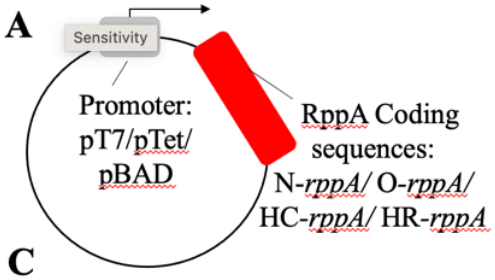
4.4.1. Design of the vector system

To design a library of constructs to improve expression, we focused on two commonly varied refactoring parameters: promoter choice and codon usage. We used four distinct coding sequences and four distinct promoters (**Figure 4.1A**). In terms of promoter choice, while the T7-*lac* promoter(209) combination is the basis of the most commonly used expression system in *E. coli* (particularly the pET expression system)(194,195) other promoters that are not strong as the T7 promoter nor as leaky as the ITPG inducible *lac* operon are sometimes used to promote soluble expression of challenging to express proteins *in vivo*(209). Other commonly used vectors include the pTet promoter, which is anhydrotetracycline inducible(209), and the pBAD promoter that is arabinose inducible(207) both of which are less leaky than the *lac* operon (**Figure 4.1B**). To obtain these constructs, we used a series of BioBrick vectors designed by Keasling and coworkers(209) that included vector pBbE2k (harboring the pTet promoter), and pBbE8k (harboring the pBAD promoter), and pBbE7k (harboring the T7 promoter under control of the *lac* operon).

To complement this series of promoters, we varied synonymous coding sequences. When expressing proteins from a high GC bacterium that differs substantially from *E. coli* in terms of codon usage, there are several approaches that can be taken. While sometimes the natural coding sequence results in successful protein expression in *E. coli*, protein expression and folding issues can occur. These issues with expression and poor folding/solubility can originate from a mismatch between tRNA pools typical of each organism, which can change the rate of translation (e.g., cause stalls at the ribosome) and potentially disrupt appropriate co-translational folding(94,210–212). Codon optimization is a common strategy to alter codon assignments, and appropriate algorithms are readily accessible from most commercial gene synthesis companies via replacement by codons used more frequently in the host's genome or transcriptome(213–219). While these genome and transcriptome-optimized sequences can aid in the successful expression of the heterologous product, improperly folded products can still result(143,220). Another approach is codon "harmonization," which has been posited to improve expression via better co-translational folding(64). Codon harmonization involves identifying and replicating patterns of codon usage in the donor organism with comparable patterns of codon usage in the heterologous host(144,221). Typically, synonymous codons (or sliding windows of these codons) are assigned computational estimates of their frequency of appearance in the original host organism. Next, single codons are changed based on frequencies estimated in the desired heterologous host to better replicate the source organism's frequency patterns, which enables stalling patterns at the ribosome that is more akin to how they originally evolved and therefore might result in proper protein folding(222).

To explore the effect of synonymous coding, four constructs were compared. The native coding sequence amplified directly from *Streptomyces griseus* genomic DNA, routine codon optimization was performed by Integrated DNA Technology's (IDT) codon optimization algorithm, and finally, two codon harmonization constructs were designed, the first using the CHARMING (for Codon HARMonizING) (HC-*rppA*)(220) and a new method based on ribosome overhead costs Stochastic Evolutionary Model of Protein Production Rate (ROC-SEMPPR)(223) (HR-*rppA*) (**Figure 4.1C**). Briefly, the CHARMING method applies a relative measure of codon usage called "% MinMax" (%MM)(220). %MM values are computed as described by Chaney et al.(212,224,225) using overall codon usage from an organism obtainable from various sources including codon usage information tabulated in the international DNA sequence database (Kazusa: https://www.kazusa.or.jp/codon/)(226). CHARMING uses a sliding window to estimate %MM-based deviations between the original and target organism codon usage for a given protein. While large deviations exist within one or more windows, single synonymous changes are made that best "harmonize" the values, i.e., reduce the overall %MM difference in that specific window. In short, this algorithm will minimize the sum of

A



B

Vector	Promoter
pBbE2k	pTet
pBbE7k	pT7
pBbE8k	pBAD
pBbEJk	pJ23101

C

Coding Strategy	Description
N- *rppA*	Native *rppA*
O- *rppA*	*rppA* Codon optimized
HC- *rppA*	*rppA* codon harmonized using the CHARMING algorithm
HR- *rppA*	*rppA* codon harmonized using the ROC-SEMPPR algorithm

Figure 4.1. Design and nomenclature of plasmids in this study.
(A) Each plasmid tested is composed of two modules: a backbone containing different promoters: pT7, pTet, pBAD and different coding sequences of the gene of interest (rppA): N-rppA, O-rppA, HC-rppA, and HR-rppA. (B) Names of the BioBrick vectors (pBb) carrying different promoters(209). (C) Names of the varied coding sequences.

| MM_original – MM_target | over all windows and will proceed until five consecutive iterations where no beneficial, i.e., reduces differences between the %MM values, changes are found. The size of the windows used was the CHARMING default value, which was set to be most consistent with ribosome fingerprint-based pausing estimates(220). In contrast, ROC-SEMPPR takes an evolutionary approach to estimating the translational efficiency of an amino acid's synonymous codons within a given organism. ROC-SEMPPR does so by fitting a probabilistic, population genetics-based model of sequence evolution, which includes the contributions of selection, mutation bias, and genetic drift, to an organism's coding sequences(223,227). By simultaneously analyzing intragenic and intergenic patterns of synonymous codon usage within a genome, ROC-SEMPPR uses estimates differences in ribosome pausing times among synonymous codons translational efficiency mutation bias between codons, and differences in protein production rates between genes. Fitting ROC-SEMPPR separately to the donor (in this case *Streptomyces griseus*) and host (in this case *E. coli*) genomes enables the ability to rank each amino acid's synonymous codons by their translational efficiencies within the donor and host, respectively.

The promoter and coding sequence combinations represent a total of 16 different constructs. The naming convention for the components of the 16 constructs is detailed in **Figure 1**.

4.4.2. Initial cell-free experiments for a type III PKS enable the production of flaviolin.

RppA catalyzes polyketide synthesis by condensing and cyclizing five molecules of the starter unit malonyl-CoA, resulting the pentaketide tetrahydroxynapthalene (THN)(204,205). Subsequently, THN undergoes a spontaneous oxidation reaction to convert THN to flaviolin (**Figure 4.2A**). The formation of the red-brown flaviolin pigment can thus be monitored as it readily absorbs light at 340 nm(206). As metabolite production in an *E. coli* lysate-based cell-free system is correlated with the amount of protein that is expressed(79), we used the amount of flaviolin produced as a proxy for estimating catalytically functional RppA production. To establish assay conditions, we first sought to determine the threshold for pigment production against a cell lysate background. To do so, we spiked purified flaviolin into lysate preparations in the absence of DNA. Sufficient pigment concentrations could be detected at the micromolar range, demonstrating sufficient sensitivity to proceed (**Appendix Figure S4.1**). With an assay established, pigment production could be tested at variable temperature conditions and plasmid DNA concentrations.

To define temperature conditions for conducting CFE experiments, we used the pET28b expression plasmid and codon-optimized *rppA* (O-*rppA*) in a lysate-based system using *E. coli* BL21 Star(DE3), a BL21(DE3) variant that has the DE3 lysogen under control

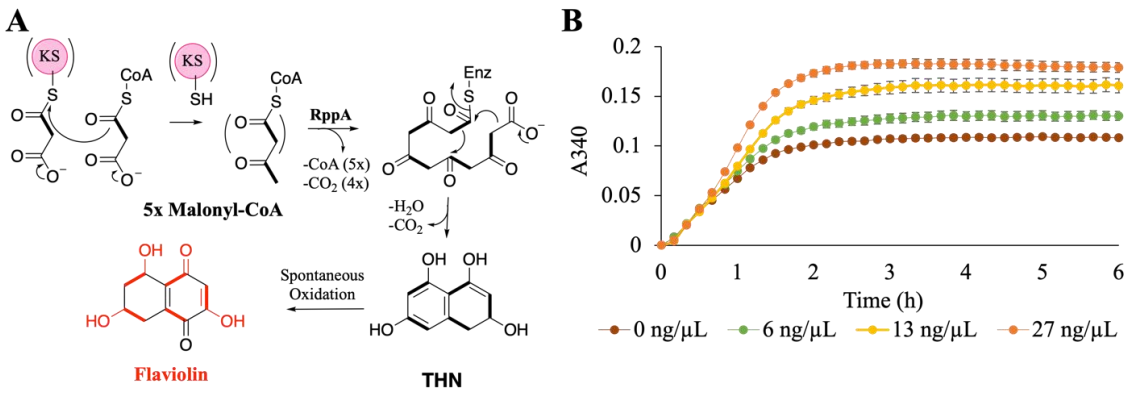


Figure 4.2. Flaviolin production via RppA in CFE reactions.
(A) Schematic of flaviolin biosynthetic pathway by type III PKS, RppA. (B) CFE reactions were initiated with increasing concentrations of codon optimized *rppA* (O-*rppA*) driven by pT7 promoter in plasmid DNA, resulting in increasing flaviolin production (observed by increase of absorbance at 340 nm).

of a *lac*UV5 promoter(228). We have previously had success using this lysate for the production of pigments from *Streptomyces*(79). Our rationale for this initial choice of coding sequence and promoter was twofold: 1) prior studies from our laboratory suggest that heterologous expression of enzymes originating from *Streptomyces* that form pigments show improved expression and solubility when using *E. coli* optimized sequences(94) and 2) the T7 promoters (and specifically pET vectors) have been widely used in CFE(79,229–233). These initial experiments revealed superior pigment production at 30˚C, so we proceeded with 30˚C for all subsequent experiments (**Appendix Figure S4.2**). To remove confounding variables from differing intergenic regions and a different origin of replication in the pET vector as compared to the BioBrick vectors, we repeated this experiment using pBbE7k as the backbone for consistency with the pBAD and pTet constructs. Increasing the amount of the plasmid construct pBbE7k-O-*rppA* (containing *E. coli* codon-optimized *rppA* driven by pT7 promoter) with *E. coli* BL21 Star(DE3) lysate containing endogenous IPTG at 30˚C demonstrated that the DNA template concentration affects protein expression and therefore product formation (**Figure 4.2B**).

4.4.3. Establishing the application of non-IPTG inducers for cell-free with other promoters

The ITPG inducible T7-*lac* expression system, especially the pET expression system is by far the most heavily used expression system for heterologous recombinant protein production in *E. coli* due to its engineered promoter strength(234). However, its extreme promoter strength, combined with the leakiness of the *lac* operon, adds to the metabolic burden of the cell resulting in decreased fitness less than optimal protein expression. For proteins that seem to be better expressed and appropriately folded with tighter transcriptional control, other inducible promoter systems have been developed, such as the arabinose inducible pBAD and anhydrotetracycline inducible pTet systems. Indeed, there is precedence for alternate promoter systems improving expression of challenging, complex proteins from *Streptomyces*; proteins from the borrelidin polyketide synthase from *Streptomyces parvulus* were found to have improved expression using the pTet promoter when compared to the T7 system(232,235,236). Extensive efforts have been made to apply the T7 promoter series under the *lac* operator in CFE(229,237). However, other inducible systems remain extremely underexplored, with only a few reports of their usage in lysate-based CFE systems(238–240).

To compare the performances of each promoter in a lysate-based system, extracts were first prepared from *E. coli* BL21 Star(DE3) in the absence of IPTG. While IPTG is typically added to BL21 Star(DE3) cultures to promote the expression of T7 RNA polymerase prior to cell lysis, we omitted this step to prepare a lysate background that is appropriate for the comparison of non-IPTG inducible promoters. Using this batch of lysate, reactions were then first optimized to express RppA under pT7 and with the

supplementation of IPTG to the lysate reaction. IPTG concentrations were supplemented to reactions containing 50 ng/µL T7 RNA polymerase(241). Maximal flaviolin production was observed with a concentration of 500 µM IPTG (**Appendix Figures S4.3A, C**). The same lysate preparations were then used for protein expression driven by the pTet and pBAD promoters under different ranges of anhydrotetracycline and L-arabinose, respectively. However, under these conditions, we did not observe detectable flaviolin production. We hypothesized that flaviolin signals are not detectable under these conditions due to lower soluble protein expression, and, consequently, less flux being driven from endogenous malonyl-CoA precursor pools to flaviolin formation. Thus, we hypothesized that increasing substrate availability would increase flaviolin production to detectable levels. To test this, malonyl-CoA was added to reactions containing pBbE7k-O-*rppA* plasmid DNA. Indeed, adding up to 500 µM malonyl-CoA to the reactions enabled higher levels of flaviolin synthesis (**Appendix Figures S4.3B, D**). *E. coli* can possess low levels of malonyl-CoA endogenously(242), therefore, with the pT7 system, there is a high level of RppA protein that flaviolin can be detectably produced even without malonyl-CoA supplementation. The low expression from the pTet/pBAD constructs combined with the low endogenous malonyl-CoA pools may be why the signals from pTet/pBAD constructs are low within the timeframe we measured. Applying this concentration (500 µM) of malonyl-CoA to systems driven by pTet and pBAD promoters results in the successful detection of flaviolin, albeit at later time points, confirming that soluble RppA expression is attainable with these promoters. A wide range of anhydrotetracycline and L-arabinose concentrations were tested using pBbE2k-O-*rppA* (O-*rppA* driven by pTet) and pBbE8k-O-*rppA* (O-*rppA* driven by pBAD), respectively (**Appendix Figures S4.4, S4.5**). For the pTet promoter, we observed the greatest expression level at 50 µM anhydrotetracycline (**Appendix Figures S4.4A, B)**, whereas 10 mM L-arabinose was greatest for the pBAD promoter (**Appendix Figures S4.5A, B)**.

Intriguingly, when optimized inducer concentrations are applied, overall flaviolin production is comparable between the pTet and pT7 conditions, even though synthesis is clearly delayed in the former system. Faster expression in the pT7 system could be due to promoter strength or the availability of exogenously supplied T7 RNA polymerase, whereas the other promoters rely on low levels of endogenous *E. coli* RNA polymerase. To confirm that fast expression from pT7 is a result of this promoter's strength, we first supplied reactions with decreasing concentrations of T7 RNA polymerase. While overall flaviolin production decreases with lower polymerase levels, flaviolin synthesis still begins within the first hour under all conditions. These data imply that fast expression under pT7 is due to this system being less tightly regulated compared to pTet and pBAD and not necessarily the availability of the polymerase (**Appendix Figures S4.6**). To further interrogate this phenomenon, we tested our promoter strategy with sfGFP, allowing us to distinguish protein expression from enzyme catalysis or precursor availability. As an initial step, we verified that inducer concentrations performed best for RppA expression were also performed best for inducing sfGFP expression under the control of different promoters in

CFE reactions (**Appendix Figures S4.7A-C**). We subsequently compared sfGFP synthesis from these inducible promoters and a constitutive promoter, pJ23101, in an analogous biobrick vector (pBbEJk). Like expression from pTet and pBAD, constitutive expression would also rely on endogenous *E. coli* RNA polymerase pools but should be quicker in theory because of a lack of regulation. Evidently, there is a delay in the expression of sfGFP under the control of either pTet, pBAD, and pJ23101 promoters compared to pT7 (**Appendix Figures S4.7D**). While the constitutive promoter produced relatively low levels of sfGFP, it did result in sfGFP expression over a faster timeframe compared to pTet and pBAD (**Appendix Figures S4.7D & E**). Thus, delayed CFE from pTet and pBAD is likely a result of their tighter regulation compared to pT7.

4.4.4. Synergistic Effect of Promoter and Coding Strategy in Refactoring Proteins for RppA CFE

After establishing a set of experimental conditions for detectable flaviolin formation driven by non-T7 inducible promoters, we sought to determine the synergistic effect of promoter plus coding strategy in refactoring a protein for optimized expression in our cell-free system. Overall, we found the strongest promoter-coding strategy to be the pBAD promoter using the ROC-SEMPPR method (HR-*rppA*) which is slightly higher than the CHARMING method (HC-*rppA*) driven by the pTet promoter and has significantly higher expression than all other constructs (**Figure 4.3A-C)**. Whether or not ROC-SEMPPR will perform similarly well in other situations remains to be determined. It does, however, suggest that naively ranking codons based on their occurrence in a genome (which ignores the role of mutation bias, variation in expression between genes) or transcriptome (which ignores the role of mutation bias and the limits drift places on adaptation), while useful, can be improved upon. In addition, we observed that expression with the pT7 promoter can be drastically improved when the lysate is supplied with endogenous IPTG as opposed to exogenous IPTG during production of the cell lysate to express the T7 RNAP in the BL21(DE3)Star strain (**Appendix Figure S4.6**). Interestingly, with the pT7 promoter, the HC construct produced the least amount of flaviolin, even lower than the natively coded sequence cloned from genomic DNA. The CHARMING method used to generate the HC construct uses sliding windows to estimate local rates of translation across the coding sequence, e.g., to best facilitate natural ribosomal stalling (see Methods)(220); however, because few rare codons are observed in the native *rppA* coding sequence, it appears that this method is not necessary to promote functional protein production using a pT7 promoter (**Figure 4.3A**). The trend of each coding sequence considered is different for each promoter. For the pTet promoter, HC-*rppA* has the best expression, then HR-*rppA* is the second best, while the native construct (N-*rppA*) does not express at all (**Figure 4.3B**). Similarly, with the pBAD promoter, the native construct has low expression while HR-*rppA* has the best expression followed by HC-*rppA* (**Figure 4.3C**). Because these codon harmonization models do not account for inducible expression,

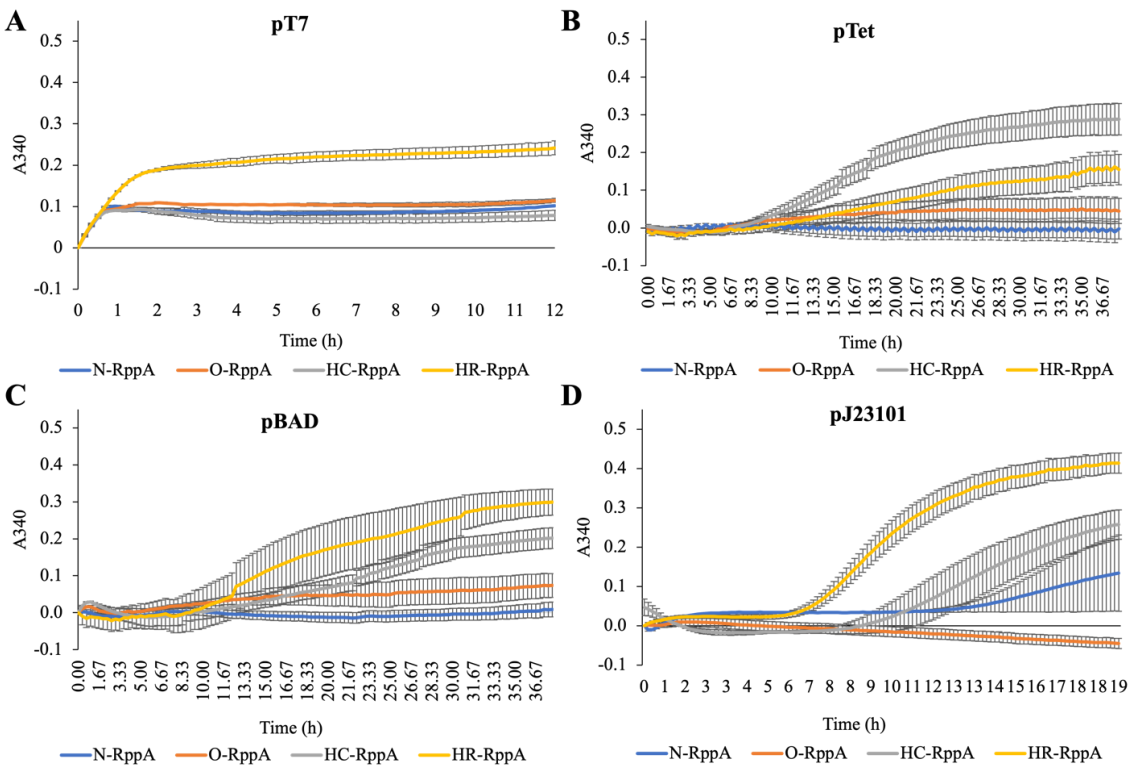


Figure 4.3. Flaviolin measurement in CFE reactions initiated with plasmids carrying different promoter and coding sequence combinations.
Error bars represent the standard error of the mean (n = 3). (A) CFE reactions containing one of four pT7 promoter constructs, 50 ng/µL T7 RNA polymerase, and 500 µM IPTG. Reactions were run in triplicate and read every 10 mins for 12 hr. (B) CFE reactions containing one of four pTet promoter constructs, 500 µM malonyl-CoA, and 50 µM tetracycline. Reactions were run in triplicate and read every 20 mins for 38 hours. (C) CFE reaction containing four pBAD promoter constructs plasmid DNA, 500 µM malonyl-CoA, and 10 mM l-arabinose. Reactions were run in triplicate and read every 20 mins for 38 hours. (D) CFE reaction containing four pJ23101 promoter constructs plasmid DNA and 500 µM malonyl-CoA. Reactions were run in triplicate and read every 10 mins for 20 hours. All reactions were prepared with E. coli BL21 Star(DE3) lysates. Error bars represent the standard error of the mean (n = 3).

we sought to determine the effects of these re-coding strategies on flaviolin production when RppA is expressed under pJ23101, allowing us to decouple expression from induction. In this case, HR-*rppA* expressed drastically better than other constructs, whereas, in contrast, the O-*rppA* did not express at all. Importantly, these experiments show that the choice of codon optimization/harmonization techniques and promoter both impact protein expression synergistically, which is an important consideration for future efforts to use CFE for high-yield protein expression. Additionally, unlike RppA, sfGFP expressed best under pT7 control while the constitutive promoter does not express well in these CFE conditions (Appendix Figure S4.7D). Thus, these results also confirm that different proteins have varying optimal CFE expression conditions(79).

4.4.5. Investigating the Utility of CFE for Prototyping Refactoring Techniques for In vivo Production

Expanding strategies for profiling expression choices to less explored choices (e.g., non-T7 promoters and lesser used refactoring strategies) and demonstrating their synergistic effects in CFE is more valuable when there are correlations between CFE and *in vivo*(230,243,244). To determine whether the refactoring strategies we explored are correlative to *in vivo* expression, we transformed each of our refactored constructs into BL21Star (DE3) cells. First, OD_{600} of the codon-optimized construct of each promoter was measured to determine inducing time. Optimal OD_{600} for each promoter/operator combination was based on literature precedent for standard ODs of induction for each promoter respectively (OD_{600} = ~0.8 for T7-*lac*, OD_{600} = ~0.2 for pBAD, OD_{600} = ~0.6 for pTet) (245–247). In a 96-well plate, from the initial culture with OD_{600} = 0.05, pT7 constructs take 3.5 hr to reach OD_{600} = 0.8, pTet constructs take 3 hours to reach OD_{600} = 0.6, and pBAD constructs take 70 mins to reach OD_{600} = 0.2. Next, we varied the inducer concentrations for each promoter (**Appendix Figure S4.8**). Of the conditions we evaluated, we found that inducer concentrations are different with CFE reactions: IPTG concentration reaches greatest expression at 500 µM, anhydrotetracycline at 1000 nM, and L-arabinose at 5 mM. These conditions correlate with previous reported for RFP, thus, they were used for sfGFP expression(209).

When comparing the effect of codon optimization/harmonization on RppA expression, without considering promoter choice, trends only correlate between the *in vivo* and *in vitro* experiments in the pT7 and the constitutive promoter data (**Figures 4.3A&D, 4.4A&D**). When expressing with pTet, the HC-*rppA* outperforms HR-*rppA in vitro* while these two harmonized sequences perform similarly *in vivo*. Differences in flaviolin synthesis from varying coding sequences under pBAD expression are indistinguishable *in*

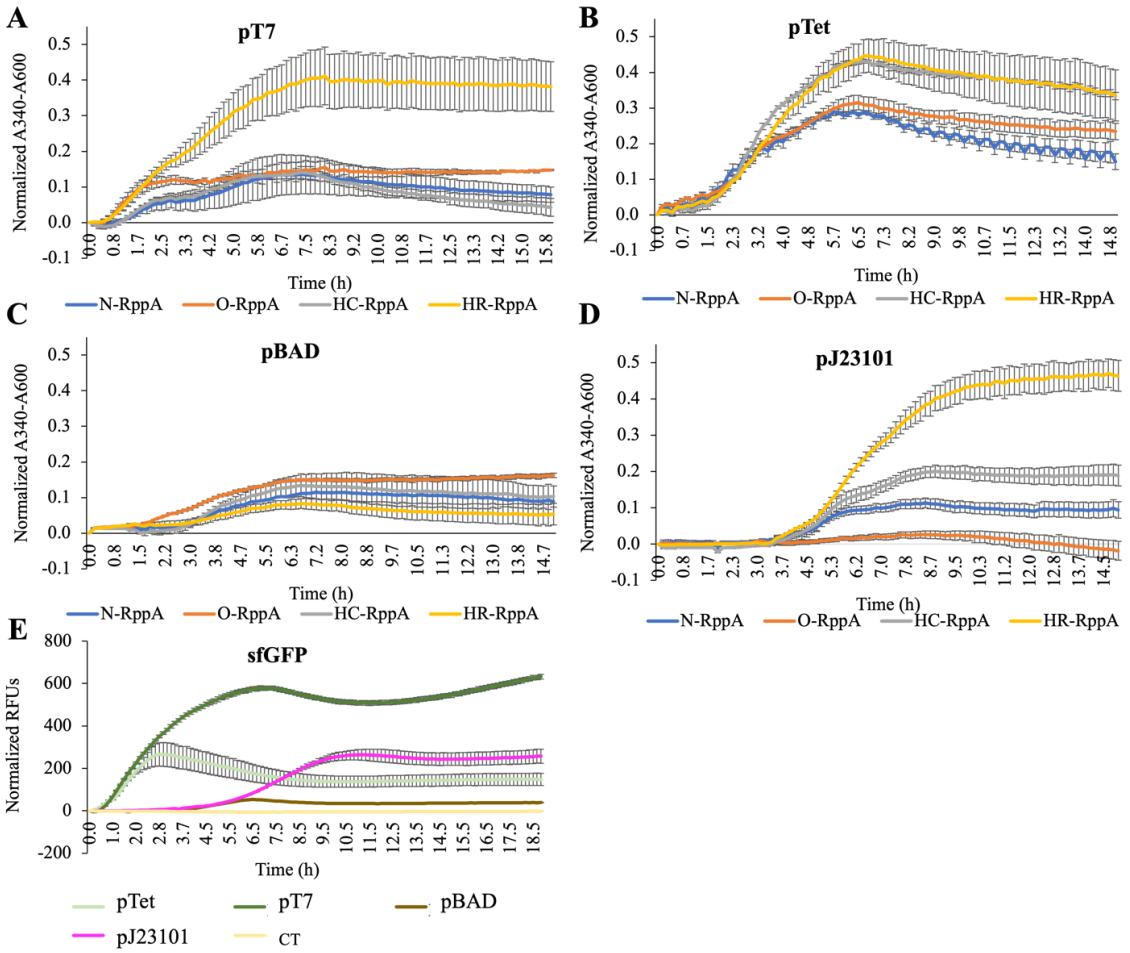


Figure 4.4. Flaviolin formation in E. coli BL21 (Star)DE3 cells carrying different promoter and coding sequence combinations.
Flaviolin production in cells carrying (A) pT7 promoter constructs post-induction with 500 µM IPTG (added after 3.5 hours of growth), (B) pTet promoter constructs post-induction with100 nM anhydrotetracycline (added after 3 hours of growth), (C) pBAD constructs post-induction with 5000 µM l-arabinose (added after 1 hour and 10 min), and (D) pJ23101 constructs. (E) Expression of BL21 Star(DE3) cells harboring sfGFP. Control (CT) is BL21 Star(DE3) without plasmid DNA. Reactions were run in triplicate and read every 10 minutes for 20 hours. Error bars represent the standard error of the mean (n = 3).

vivo (**Figure 4.4C**), and generally lower compared to production in the cell-free system (**Figure 4.3C**). Efficient catabolism of L-arabinose by *E. coli* cells is a drawback of arabinose-inducible promoters, which may be a potential cause for better flaviolin synthesis in pBAD-regulated CFE(207). Notably, the current CFE system cannot be used to prototype promoter choice for the *in vivo* expression of RppA, given any coding sequence (**Figure 4.3 and Figure 4.4**). The pT7 expression measurements were collected when the inducible IPTG was added, which was after 3.5 hours of growth (**Figure 4.4A**). While the measurement of the pJ23101 constructs were collected right after inoculation into the 96-well plate. It takes about 3-4 hours for the cells to grow before they start to produce pigment. This explains the delay occurring in the pJ23101 promoter (**Figure 4.4D**). This is also true for sfGFP expression. While sfGFP expressed highest using pT7 in both systems, the pBAD-sfGFP expressed better in CFE while the constitutive promoter and pTet promoter are expressed better *in vivo* (**Appendix Figures S4.7D&4.4E**). From this set of experiments, this suggests that there is strong correlation, but not complete alignment between expressing in cell free vs. *in vivo*.

4.5. Conclusions

Natural product synthesis in non-native contexts requires the successful translation and folding of biosynthetic genes. Refactoring choices to improve the heterologous expression and activity of these enzymes include the use of inducible or constitutive promoters and codon optimization/harmonization strategies. These elements are commonly explored for *in vivo* protein synthesis purposes, but long Design-Build-Test-Learn (DBTL) cycles associated with cellular engineering can be a limiter when evaluating gene refactoring strategies. Cell-free systems enable the accelerated testing of such tactics and thus enable the rapid optimization of refactoring choices for *in vitro* or *in vivo* expression. However, besides pT7 promoter systems and a selection of constitutive promoters(248), other promoter systems have not been used extensively in CFE. Tools for codon harmonization, as opposed to codon optimization, have also not been considered for CFE. Thus, we aimed to explore whether inducible promoter systems and codon harmonization can benefit *in vitro* protein synthesis or be prototyped in CFE reactions for *in vivo* implementation.

In the cell-free context, we show that inducible pTet- and pBAD-regulated expression, while slower than pT7, allow higher yields of flaviolin. The same is true for constitutive expression as codon harmonization algorithms are likely to more accurately measure the “tempo” of translation elongation(220,249). We demonstrate that even in a gene with mostly efficient codons (74.5% rank 1, 21.7% rank 2; avg rank 1.33 based on ROC-SEMPPR), codon harmonization improves natural product synthesis more than uniform codon optimization across the cell-free and cellular environments considered here (7/8 cases). This is consistent with Keasling and coworkers recent report that in some, but

not all heterologous hosts, codon harmonization can be superior to other codon optimization methods to express a type I polyketide synthase gene from actinomycete origin(193), providing further support to the notion that codon harmonization should be explored more generally to promote improved protein production from biosynthetic genes from Actinobacteria. Interestingly, we found that different harmonization methods do not work equally well for *rppA*. Consistent with the original protein coding sequence having few "slow" codons, which probably affect co-translational folding the most(212), in 6 out of the 8 cases evolutionarily based harmonization, i.e., fitting the evolutionarily based ROC-SEMPPR to the donor and host genomes to determine and replace based on individual codon ranks, performed substantially better than the window-based CHARMING approach (in one of the two other cases, the differences were almost indistinguishable). We also found that the choice of promoters influences the outcome of refactoring coding sequences. These interactions also vary between *in vitro* and *in vivo* reactions, particularly for non-pT7 inducible promoters, for which the relative activities of promoters and the synergies between promoter usage and coding sequences poorly correlate between cell-free and cell-based systems. In conclusion, refactoring promoters and/or coding sequences via CFE can be a valuable strategy to rapidly screen for catalytically functional production of enzymes from BCGs. This can in turn accelerate DBTL cycles to generate valuable metabolites.

Appendix

Table S4.1. Genes sequences were used in this study; highlighted nucleotides are for the strep tag.

Name	Sequence
Native *rppA*	atggcgaccctgtgccgaccggccatcgctgtgcccgagcacgtcatcacgatgcagcagaccctg gacctggcccgggagacccatgccgggcacccgcagcgcgacctcgtcctgaggctcatccagaa caccggcgtccagacccggcacctcgtgcagcccatcgagaagaccctggcgcaccccggattcg aggtgcgcaaccaggtgtacgaggccgaggccaagacccgggtccccgaagtcgtccggcgggc gctcgccaacgccgagaccgagccgtccgagatcgacctgatcgtctacgtctcctgcacgggtttca tgatgccctcgctgaccgcgtggatcatcaacagcatgggcttccggcccgagacccgccaactgcc catcgcccagctcggctgtgcggcgggcggcgcggcgatcaaccgcgcgcacgacttctgcgtgg cctaccccgactccaacgtcctcatcgtgtcctgcgagttctgctcgctgtgctaccagcccaccgaca tcggggtcggttccctgctctccaacggactcttcggcgacgcgctctccgcggccgtcgtacgggg acagggcggcaccggcatgcgcctggagcgcaacggctcccacctggtgcccgacaccgaggact ggatctcctacgcggtccgcgacaccgggttccacttccagctggacaagcgggtcccgggcaccat ggagatgctcgccccggtgctcctggacctggtcgacctgcacggctggtccgtcccgaacatggac ttcttcatcgtccacgcgggcggaccgcgcatcctggacgacctctgccacttcctcgacctgccgcc cgagatgttccgctacagccgggccaccctcaccgaacgcggcaacatcgcgagctccgtcgtcttc gacgcgctggcgcgcctcttcgacgacggcggcgccgccgagtccgcgcaggggctcatcgccg gcttcggtcccggcatcaccgccgaggtggccgtggggagttgggccaaggaaggcctcgggggcg gacgtcggacgcgacctcgacgagctggagctgaccgccggcgttgcgctgtccggctggagcca cccgcagttcgaaaaataa
E. coli codon optimized *rppA*	atggctactctgtgtcgccctgcgattgccgttccagagcacgttatcacgatgcaacaaactttggatct tgcccgtgagacccacgcaggccatcctcaacgtgacctggtgttacgtttgatccagaatactggagt tcaaacccgccatttagttcagccaatcgaaaagacattggcacatcctgggtttgaagttcgcaacca agtctatgaagctgaagcaaaaacgcgcgtacctgaagtggttcgccgtgctctggcgaatgcggaa acggagcccagtgaaatcgatttaattgtgtatgtttcttgcaccggttttatgatgccgtctcttacggcat ggattatcaatagtatgggctttcgtcctgagacacgccaactgccaatcgcacaattaggatgtgcag cgggaggtgcggcgattaaccgcgcccatgatttttgcgtagcataccccgactccaatgttttaattgt gagctgtgagttttgctcgctgtgttatcaaccaacggatattggcgtgggttccttgttgtccaacggatt gtttggggacgcccttagtgcggctgttgtgcgtggtcaggggggtaccggcatgcgccttgaacgca atgggtcccacttggttcccgacacggaggattggatctcatacgcagtgcgtgatacaggcttccattt ccaattagacaaacgcgtacctgggacgatggaaatgttagcgccggtactgttggacctggtggatc tgcacggttggagcgtgccaaatatggacttttttatcgtgcacgccggtggacctcgcatccttgatga cttatgtcacttttggatctgcctccagagatgttccgctactctcgcgccactcttacggaacgcggca atattgcaagctcagtcgtatttgatgcattggcgcgcttatttgacgacggaggcgcagccgagtccg cccaagggctgatcgcggggtttggtccggggatcacggcggaagtagcagtcggtagctgggcca aggaggggttaggagcggacgttgggcgtgatctggatgagttggaactgaccgccggagttgctct gagtggatggagccacccgcagttcgaaaaataa

Table S4.1. Continued.

rppA codon harmonized using the CHARMING algorithm	atggccaccctgtgcagaccggcgattgctgttccagaacatgtgattacgatgcagcagaccctgga tctggcccgtgaaacccacgcgggtcatccgcagcgcgatttagtgctgcgtctgattcagaacaccg gcgtgcagacccgtcatctggttcagccaattgaaaaaaccctggcccatccaggatttgaagtgcgc aaccaggtttatgaagcggaagcgaaaacccgcgtgccagaggtggtgcgccgcgcgctggccaa cgcggaaaccgaaccgagcgaaattgatctgattgtgtatgtgagctgcacggggtttatgatgccaa gcctgaccgcctggattattaactccatgggctttcgtccggaaacccgccaactgccgattgcccagc tgggctgtgccgccggtggtgccgccattaaccgcgcccatgatttttgcgttgcgtatccggatagca acgtgctgattgttagttgcgaattttgcagtctgtgctatcagccaaccgatattggtgtggggagcctg ttaagcaacggtttatttggcgatgcgctgagcgccgcggtggtacgtggacagggcggcaccggca tgcgcctggaacgcaacggcagccatctggtgccagataccgaagattggattagctatgccgttcgc gataccggttttcactttcagctggataaacgtgtgccgggcaccatggaaatgctggccccggtgctg ctggacctggtggatctgcatggctggagcgtgccgaacatggatttttttattgtgcatgccggcggac cgcgcattctggatgatttatgccattttctggatctgccgccagaaatgtttcgctattcccgtgcgaccc tgaccgagcgcggcaacattgcctccagcgtggtgtttgatgcactggcacgtctgttcgatgatggcg gcgcggcggaaagcgcccagggcttaattgcgggctttggtccaggcattaccgcggaagttgcggt tggttcgtgggcgaaagaaggcttaggtgccgatgtgggacgcgatctggatgaactggaactgacg gcggggcgtagccctgagcggctggagccacccgcagttcgaaaaatag
rppA codon harmonized using the ROC-SEMPPR algorithm	atggcaactctctgccggcctgctatcgccgtgccggaacacgttatcaccatgcagcagactctcga cctcgctcgcgaaactcatgctgggcaccctcagcgtgacctggttctccgactgatccagaacactg gtgttcagactcgccacctggtgcagccgatcgaaaaaactctcgcacacccgggcttcgaagtgcgt aaccaggtgtacgaagctgaagctaaaactcgcgttccggaggttgttcgccgcgcactggctaacg ctgaaactgaaccttctgaaatcgacctcatcgtttacgtttcttgcaccggattcatgatgccgtccctca ctgcatggatcatcaacagcatgggtttccgcccggaaactcgtcaactcccgatcgctcagctgggtt gtgcagcaggtggtgcagcaatcaaccgtgcacacgacttctgcgtggcttacccggactctaacgtt ctgatcgtgtcttgcgaattctgctccctctgctaccagccgactgacatcggggttggatctctcctgtct aacggcctgttcggtgacgcactgtctgcagctgttgtacgcggccagggtggtactggtatgcgtctc gaacgtaacggttctcacctcgtgccggacactgaagactggatctcttacgcagttcgtgacactggg ttccacttccagctcgacaaacgcgttcctggtactatggaaatgctggctcctgtgctgctcgacctcgt tgacctccacggttggtctgttcctaacatggacttcttcatcgttcacgcaggtggccctcgtatcctcg acgacctgtgccacttcctggacctccctccggaaatgttccgttacagccgcgctactctgactgagc gtggtaacatcgcaagctctgttgttttcgacgcactcgcacgtctgttcgacgacggtggtgctgctga atctgcacaggggctgatcgctggtttcggaccgggtatcactgctgaagtggctgtggggagttggg ctaaagagggtctgggggcagacgttggccgtgacctggacgaactcgaactcactgctggtgtcgc actctctggttggagccacccgcagttcgaaaaataa
*sf*GFP	atgagcaaaggtgaagaactgtttaccggcgttgtgccgattctggtggaactggatggcgatgtgaa cggtcacaaattcagcgtgcgtggtgaaggtgaaggcgatgccacgattggcaaactgacgctgaaa tttatctgcaccaccggcaaactgccggtgccgtggccgacgctggtgaccaccctgacctatggcgt tcagtgttttagtcgctatccggatcacatgaaacgtcacgatttctttaaatctgcaatgccggaaggct atgtgcaggaacgtacgattagctttaaagatgatggcaaatataaaacgcgcgccgttgtgaaatttga aggcgataccctggtgaaccgcattgaactgaaaggcacggattttaaagaagatggcaatatcctgg

Table S4.1. Continued.

*sf*GFP (continued)	gccataaactggaatacaactttaatagccataatgtttatattacggcggataaacagaaaaatggcat caaagcgaattttaccgttcgccataacgttgaagatggcagtgtgcagctggcagatcattatcagca gaataccccgattggtgatggtccggtgctgctgccggataatcattatctgagcacgcagaccgttct gtctaaagatccgaacgaaaaaggcacgcgggaccacatggttctgcacgaatatgtgaatgcggca ggtattacgtggagccatccgcagttcgaaaaataa

Table S4.2. Plasmid sequences were used in this study; highlighted nucleotides are for the promoters.

Name	Sequence
pBbE2k-containing pTet promoter	ggatccaaactcgagtaaggatctccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctt tcgttttatctgttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgc gtttatacctagggcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccac agaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaa aaggccgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaag tcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgc tctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctca tagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccc cgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatc gccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttg aagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttacc ttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgca agcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctc agtggaacgaaaactcacgttaagggattttggtcatgactagtgcttggattctcaccaataaaaaacgcc cggcggcaaccgagcgttctgaacaaatccagatggagttctgaggtcattactggatctatcaacaggag tccaagcgagctctcgaaccccagagtcccgctcagaagaactcgtcaagaaggcgatagaaggcgatg cgctgcgaatcgggagcggcgataccgtaaagcacgaggaagcggtcagcccattcgccgccaagctc ttcagcaatatcacgggtagccaacgctatgtcctgatagcggtccgccacacccagccggccacagtcg atgaatccagaaaagcggccattttccaccatgatattcggcaagcaggcatcgccatgggtcacgacga gatcctcgccgtcgggcatgcgcgccttgagcctggcgaacagttcggctggcgcgagcccctgatgct cttcgtccagatcatcctgatcgacaagaccggcttccatccgagtacgtgctcgctcgatgcgatgtttcgc ttggtggtcgaatgggcaggtagccggatcaagcgtatgcagccgccgcattgcatcagccatgatggat actttctcggcaggagcaaggtgagatgacaggagatcctgccccggcacttcgcccaatagcagccagt cccttcccgcttcagtgacaacgtcgagcacagctgcgcaaggaacgcccgtcgtggccagccacgata gccgcgctgcctcgtcctgcagttcattcagggcaccggacaggtcggtcttgacaaaaagaaccgggc gcccctgcgctgacagccggaacacggcggcatcagagcagccgattgtctgttgtgcccagtcatagc cgaatagcctctccacccaagcggccggagaacctgcgtgcaatccatcttgttcaatcatgcgaaacgat cctcatcctgtctcttgatcagatcatgatcccctgcgccatcagatccttggcggcaagaaagccatccagt ttactttgcagggcttcccaaccttaccagagggcgccccagctggcaattccgacgtcttaagacccactt tcacatttaagttgtttttctaatccgcatatgatcaattcaaggccgaataagaaggctggctctgcaccttgg tgatcaaataattcgatagcttgtcgtaataatggcggcatactatcagtagtaggtgtttccctttcttctttagc gacttgatgctcttgatcttccaatacgcaacctaaagtaaaatgccccacagcgctgagtgcatataatgca ttctctagtgaaaaaccttgttggcataaaaaggctaattgattttcgagagtttcatactgtttttctgtaggcc gtgtacctaaatgtacttttgctccatcgcgatgacttagtaaagcacatctaaaacttttagcgttattacgtaa aaaatcttgccagctttcccccttctaaagggcaaaagtgagtatggtgcctatctaacatctcaatggctaag gcgtcgagcaaagcccgcttatttttacatgccaatacaatgtaggctgctctacacctagcttctgggcga gtttacgggttgttaaaccttcgattccgacctcattaagcagctctaatgcgctgttaatcactttacttttatct aatctagacatcattaattcctaatttttgttgacactctatcgttgatagagttattttaccactccctatcagtga tagagaaaagaattcaaaagatcttttaagaaggagatatacat

Table S4.2. Continued.

pBbE7k-containing pT7 promoter	ggatccaaactcgagtaaggatctccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcg ttttatctgttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcgtttat acctagggcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcag gggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgtt gctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcga aacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccct gccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtat ctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcg ccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggt aacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctaca ctagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatc cggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaggat ctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggt catgactagtgcttggattctcaccaataaaaaacgcccggcggcaaccgagcgttctgaacaaatccagatg gagttctgaggtcattactggatctatcaacaggagtccaagcgagctctcgaaccccagagtcccgctcaga agaactcgtcaagaaggcgatagaaggcgatgcgctgcgaatcgggagcggcgataccgtaaagcacgag gaagcggtcagcccattcgccgccaagctcttcagcaatatcacgggtagccaacgctatgtcctgatagcgg tccgccacacccagccggccacagtcgatgaatccagaaaagcggccattttccaccatgatattcggcaagc aggcatcgccatgggtcacgacgagatcctcgccgtcgggcatgcgcgccttgagcctggcgaacagttcg gctggcgcgagcccctgatgctcttcgtccagatcatcctgatcgacaagaccggcttccatccgagtacgtgc tcgctcgatgcgatgtttcgcttggtggtcgaatgggcaggtagccggatcaagcgtatgcagccgccgcattg catcagccatgatggatactttctcggcaggagcaaggtgagatgacaggagatcctgccccggcacttcgcc caatagcagccagtcccttcccgcttcagtgacaacgtcgagcacagctgcgcaaggaacgcccgtcgtggc cagccacgatagccgcgctgcctcgtcctgcagttcattcagggcaccggacaggtcggtcttgacaaaaag aaccgggcgcccctgcgctgacagccggaacacggcggcatcagagcagccgattgtctgttgtgcccagt catagccgaatagcctctccacccaagcggccggagaacctgcgtgcaatccatcttgttcaatcatgcgaaa cgatcctcatcctgtctcttgatcagatcatgatcccctgcgccatcagatccttggcggcaagaaagccatcca gtttactttgcagggcttcccaaccttaccagagggcgccccagctggcaattccgacgtcctcactgcccgctt tccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtat tgggcgccagggtggtttttctttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctga gagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcg ggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcagcccgga ctcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgcc ctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctga atttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccg ctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatggga gaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagctt ccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagatt gtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgat

Table S4.2. Continued.

pBbE7k-containing pT7 promoter (continued)	cggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaac gccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaattcagctccgcca tcgccgcttccactttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtc tgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattga ctctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgac gctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgc cgcaaggaatggtgcatgcaaggagatggcgcccaacagtccccggccacggggcctgccaccata cccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggc gatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagagg atcgagatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattt cagaattcaaaagatcttttaagaaggagatatacat
pBbE8k-containing pBAD promoter	ggatccaaactcgagtaaggatctccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcct ttcgttttatctgttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttct gcgtttatacctagggcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatc cacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccg taaaaaggccgcgttgctggcgtttttccataggctccgccccccctgacgagcatcacaaaaatcgacgc tcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctc gtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggc gctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcac gaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagac acgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgcta cagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctga agccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtg gttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacg gggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgactagtgcttggattctcacc aataaaaaacgcccggcggcaaccgagcgttctgaacaaatccagatggagttctgaggtcattactgg atctatcaacaggagtccaagcgagctctcgaaccccagagtcccgctcagaagaactcgtcaagaag gcgatagaaggcgatgcgctgcgaatcgggagcggcgataccgtaaagcacgaggaagcggtcagc ccattcgccgccaagctcttcagcaatatcacgggtagccaacgctatgtcctgatagcggtccgccaca cccagccggccacagtcgatgaatccagaaaagcggccattttccaccatgatattcggcaagcaggca tcgccatgggtcacgacgagatcctcgccgtcgggcatgcgcgccttgagcctggcgaacagttcggct ggcgcgagcccctgatgctcttcgtccagatcatcctgatcgacaagaccggcttccatccgagtacgtg ctcgctcgatgcgatgtttcgcttggtggtcgaatgggcaggtagccggatcaagcgtatgcagccgccg cattgcatcagccatgatggatactttctcggcaggagcaaggtgagatgacaggagatcctgccccgg cacttcgcccaatagcagccagtcccttcccgcttcagtgacaacgtcgagcacagctgcgcaaggaac gcccgtcgtggccagccacgatagccgcgctgcctcgtcctgcagttcattcagggcaccggacaggtc ggtcttgacaaaaagaaccgggcgcccctgcgctgacagccggaacacggcggcatcagagcagcc gattgtctgttgtgcccagtcatagccgaatagcctctccacccaagcggccggagaacctgcgtgcaat ccatcttgttcaatcatgcgaaacgatcctcatcctgtctcttgatcagatcatgatcccctgcgccatcagat ccttggcggcaagaaagccatccagtttactttgcagggcttcccaaccttaccagagggcgccccagct

Table S4.2. Continued.

pBbE8k-containing pBAD promoter (continued)	ggcaattccgacgtcttatgacaacttgacggctacatcattcacttttttcttcacaaccggcacggaac tcgctcgggctggccccggtgcatttttttaaataccgcgagaaatagagttgatcgtcaaaaccaac attgcgaccgacggtggcgataggcatccgggtggtgctcaaaagcagcttcgcctggctgatacgt tggtcctcgcgccagcttaagacgctaatccctaactgctggcggaaaagatgtgacagacgcgac ggcgacaagcaaacatgctgtgcgacgctggcgatatcaaaattgctgtctgccaggtgatcgctga tgtactgacaagcctcgcgtacccgattatccatcggtggatggagcgactcgttaatcgcttccatgc gccgcagtaacaattgctcaagcagatttatcgccagcagctccgaatagcgcccttcccttgcccg gcgttaatgatttgcccaaacaggtcgctgaaatgcggctggtgcgcttcatccgggcgaaagaacc ccgtattggcaaatattgacggccagttaagccattcatgccagtaggcgcgcggacgaaagtaaac ccactggtgataccattcgcgagcctccggatgacgaccgtagtgatgaatctctcctggcgggaac agcaaaatatcacccggtcggcaaacaaattctcgtccctgatttttcaccacccccctgaccgcgaat ggtgagattgagaatataacctttcattcccagcggtcggtcgataaaaaaaatcgagataaccgttgg cctcaatcggcgttaaacccgccaccagatgggcattaaacgagtatcccggcagcaggggatcatt ttgcgcttcagccatacttttcatactcccgccattcagagaagaaaccaattgtccatattgcatcaga cattgccgtcactgcgtcttttactggctcttctcgctaaccaaaccggtaaccccgcttattaaaagcat tctgtaacaaagcgggaccaaagccatgacaaaaacgcgtaacaaaagtgtctataatcacggcag aaaagtccacattgattatttgcacggcgtcacactttgctatgccatagcatttttatccataagattagc ggattctacctgacgcttttatcgcaactctctactgtttctccatacccgtttttttgggaattcaaaagat cttttaagaaggagatatacat
pBbEJ23101k-containing pJ23101 promoter	tggagccacccgcagttcgaaaaataaggatccaaactcgagtaaggatctccaggcatcaaataaa acgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctctactaga gtcacactggctcaccttcgggtgggcctttctgcgtttatacctagggcgttcggctgcggcgagcg gtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaac atgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccat aggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgac aggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgc cgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctatagctcacgctgtag gtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccg accgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactg gcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagt ggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttacct tcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttg caagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctg acgctcagtggaacgaaaactcacgttaagggattttggtcatgactagtgcttggattctcaccaata aaaaacgcccggcggcaaccgagcgttctgaacaaatccagatggagttctgaggtcattactggat ctatcaacaggagtccaagcgagctctcgaaccccagagtcccgctcagaagaactcgtcaagaag gcgatagaaggcgatgcgctgcgaatcgggagcggcgataccgtaaagcacgaggaagcggtca gcccattcgccgccaagctcttcagcaatatcacgggtagccaacgctatgtcctgatagcggtccg ccacacccagccggccacagtcgatgaatccagaaaagcggccattttccaccatgatattcggcaa gcaggcatcgccatgggtcacgacgagatcctcgccgtcgggcatgcgcgccttgagcctggcga

Table S4.2. Continued.

pBbEJ23101k-containing pJ23101 promoter (continued)	acagttcggctggcgcgagcccctgatgctcttcgtccagatcatcctgatcgacaagaccggcttcc atccgagtacgtgctcgctcgatgcgatgtttcgcttggtggtcgaatgggcaggtagccggatcaag cgtatgcagccgccgcattgcatcagccatgatggatactttctcggcaggagcaaggtgagatgac aggagatcctgccccggcacttcgcccaatagcagccagtcccttcccgcttcagtgacaacgtcga gcacagctgcgcaaggaacgcccgtcgtggccagccacgatagccgcgctgcctcgtcctgcagt tcattcagggcaccggacaggtcggtcttgacaaaaagaaccgggcgcccctgcgctgacagccg gaacacggcggcatcagagcagccgattgtctgttgtgcccagtcatagccgaatagcctctccacc caagcggccggagaacctgcgtgcaatccatcttgttcaatcatgcgaaacgatcctcatcctgtctct tgatcagatcatgatcccctgcgccatcagatccttggcggcaagaaagccaccagtttactttgcag ggcttcccaaccttaccagagggcgccccagctggcaattccgacgtctttacagctagctcagtcct aggtattatgctagcaagaattcaaaagatcttttaagaaggagatatacat
pET28b-containing pT7 promoter	caccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgct gagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggagg aactatatccggattggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtg gttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcct ttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttag tgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccct gatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactgga acaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaa aaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggc acttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctca tgaattaattcttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaata ccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggca agatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaa aaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagttta tgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaacca aaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattac aaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatc aggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatca ggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccat ctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttc ccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaat cagcatccatgttggaatttaatcgcggcctagagcaagacgtttcccgttgaatatggctcataacac cccttgtattactgtttatgtaagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgtt ccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttttctgcgcgtaatc tgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaac tctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtag ttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtg gctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataagg cgcagcggtcgggctgaacgggggggttcgtgcacacagcccagcttggagcgaacgacctacac

Table S4.2. Continued.

pET28b-containing pT7 promoter (continued)	cgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggac aggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgc ctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcagggg ggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctc acatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctc gccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcgg tattttctccttacgcatctgtgcggtatttcacaccgcatatatggtgcactctcagtacaatctgctctgatg ccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccg ccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgacc gtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaa agctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttct ccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactga tgcctccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacg atacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggat gcggcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttcc acagggtagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcg tttccagactttacgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagc agcagtcgcttcacgttcgctcgcgtatcggtgattcattctgctaaccagtaaggcaaccccgccagcct agccgggtcctcaacgacaggagcacgatcatgcgcacccgtggggccgccatgccggcgataatgg cctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattc cgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgac ccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgat agtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatc ccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacc tgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagg gtggtttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgc agcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatat aacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactc ggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgc cctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcgg ctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaat gggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtacc gtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacat tagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgac gcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacacc accacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcag ggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggtt gggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctg gttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggt ttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccatt

Table S4.2. Continued.

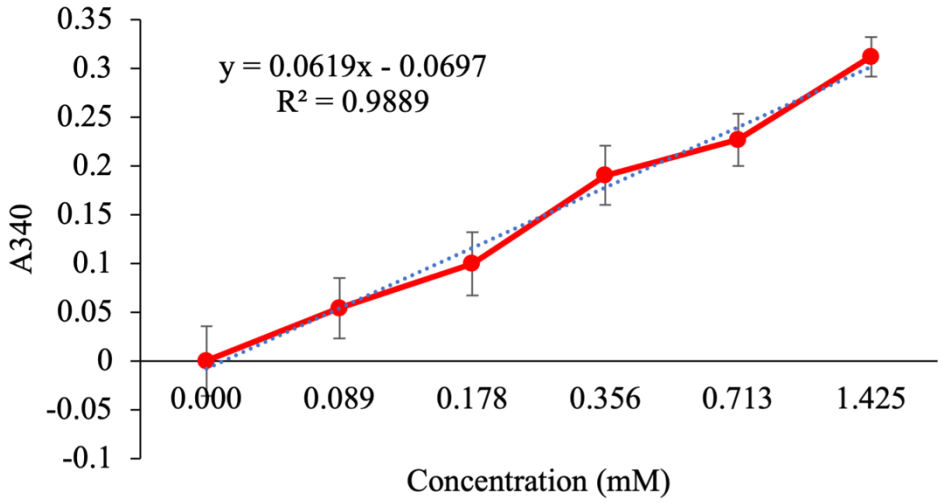
pET28b-containing pT7 promoter (continued)	cgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggtt gaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtccccc ggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccg atcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggc cacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactatagggaat tgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatataccatgggcagca gccatcatcatcatcatcacagcagcggcctggtgccgcgcggcagc

Table S4.3. Primers used in this study.
Primers #1-10 are to clone 12 constructs including four *rppA* coding sequence (native, *E. coli* codon optimized, HC and HR codon harmonized) into three plasmid backbones of pBbE2k, pBbE7k, and pBbE8k. Primers #11-14 are to clone 4 construct including four *rppA* coding sequence (native, *E. coli* codon optimized, HC and HR codon harmonized) into one plasmid backbones of pBbEJ23101k. Primers #15-18 are to clone *E. coli* codon optimized *rppA* plasmid backbones of pET28b.

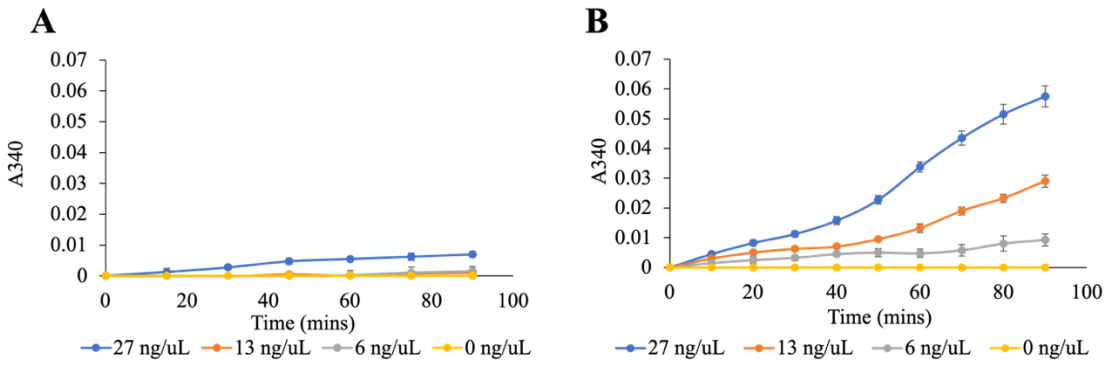
#	Primer description	Sequence
1	Native *rppA* Forward	CAAAAGATCTTTTAAGAAGGAGATATACATATGGC GACCCTGTGCCGACC
2	Native *rppA* Reverse	CCTGGAGATCCTTACTCGAGTTTGGATCCTTATTTT TCGAACTGCGGGTGGCTCCAGCCGGACAGCGCAAC GCCGGCG
3	Optimized *rppA* Forward	CAAAAGATCTTTTAAGAAGGAGATATACATATGGC TACTCTGTGTCGCCCTGC
4	Optimized *rppA* Reverse	CCTGGAGATCCTTACTCGAGTTTGGATCCTTATTTT TCGAACTGCGGGTGGCTCCATCCACTCAGAGCAAC TCCGGCGGT
5	HC *rppA* Forward	CAAAAGATCTTTTAAGAAGGAGATATACATATGGC CACCCTGTGCAGACCG
6	HC *rppA* Reverse	CCTGGAGATCCTTACTCGAGTTTGGATCCCTATTTT TCGAACTGCGGGTGGCTCCAGCCGCTCAGGGCTAC GCCCGCCG
7	HR *rppA* Forward	CAAAAGATCTTTTAAGAAGGAGATATACATATGGC AACTCTCTGCCGGCCT
8	HR *rppA* Reverse	CCTGGAGATCCTTACTCGAGTTTGGATCCTTATTTT TCGAACTGCGGGTGGCTCCAACCAGAGAGTGCGAC ACCAG
9	pBb backbone Forward	GGATCCAAACTCGAGTAAGGATCTCC
10	pBbJ backbone Reverse	GCCATATGTATATCTCCTTCTTAAAAGATCTTTTGA ATTC
11	*rppA* Forward	CGTCTTTACAGCTAGCTCAGTCCTAGGTATTATGCT AGCAAGAATTCAAAAGATCTTTTAAGAA
12	*rppA* Reverse	GCCTGGAGATCCTTACTCGAGTTTGGATCCTTATTT TTCGAACTGCGGGTGGC
13	pBbJ Forward	GGATCCAAACTCGAGTAAGGATCTCCAGGC
14	pBb Reverse	GCATAATACCTAGGACTGAGCTAGCTGTAAAGACG TCGGAATTGCCAGCTGGG

Table S4.3. Continued.

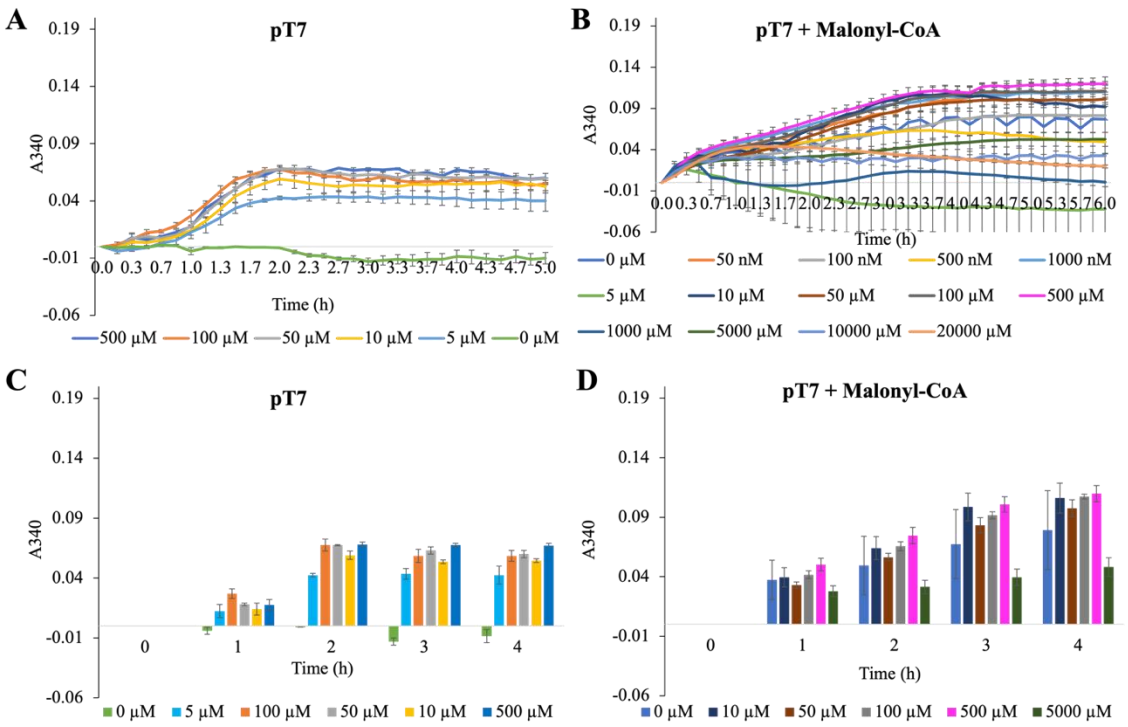
15	Optimized *rppA* Forward	CACACCAGGTCTCACAGCATGGCTACTCTGTGTCGCCCTGC
16	Optimized *rppA* Reverse	CACACCAGGTCTCATGGTGTTATCCACTCAGAGCAACTCCGGCGG
17	pET28b Forward	CACACCAGGTCTCAACCACCACTGAGATCCGGCTGCTAAC
18	pET28b Reverse	CACACCAGGTCTCAGCTGCCGCGCGGCACCAG



Appendix Figure S4.1. Standard curve of flaviolin generated by spiking increasing concentrations of purified flaviolin into BL21 Star (DE3) lysate CFE mock reactions. Absorbance measurements were taken at 340 nm. For visualization of trends, values were averaged and plotted with error bars representing the standard error of the mean (n=3).

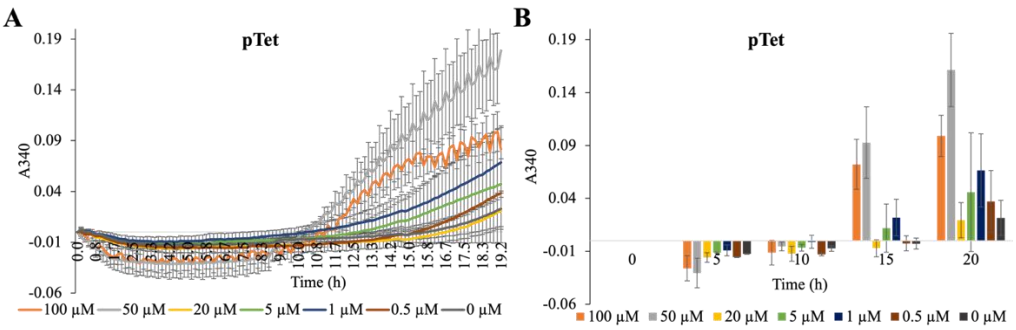


Appendix Figure S4.2. Initial E. coli lysate-based CFE reactions.
Reactions were initiated with increasing concentrations of plasmid DNA at different temperatures to determine the optimal concentration of DNA as well as temperature. (A) CFE reactions at 25˚C. (B) CFE reactions at 30˚C. All reactions were prepared with BL21 Star (DE3) extracts that contained endogenous IPTG and using plasmid DNA of optimized-*rppA* in a pET28b vector. Reactions were run in triplicate and read every 10 mins for 90 mins. Error bars represent the standard error of the mean (n = 3).

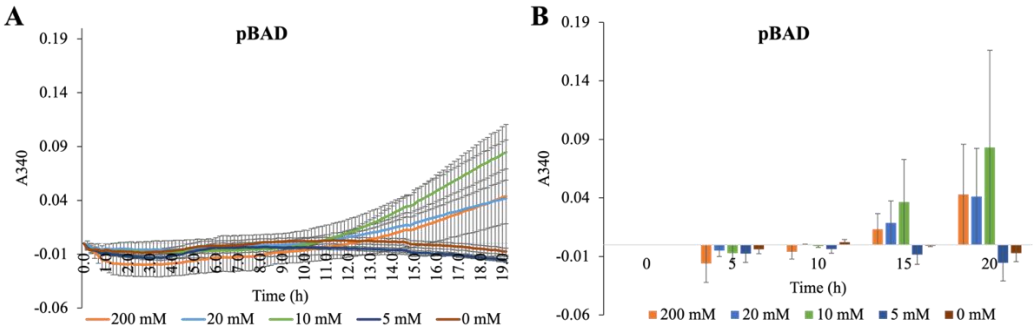


Appendix Figure S4.3. Optimization of inducer concentrations in *E. coli* lysate-based CFE reactions for RppA.

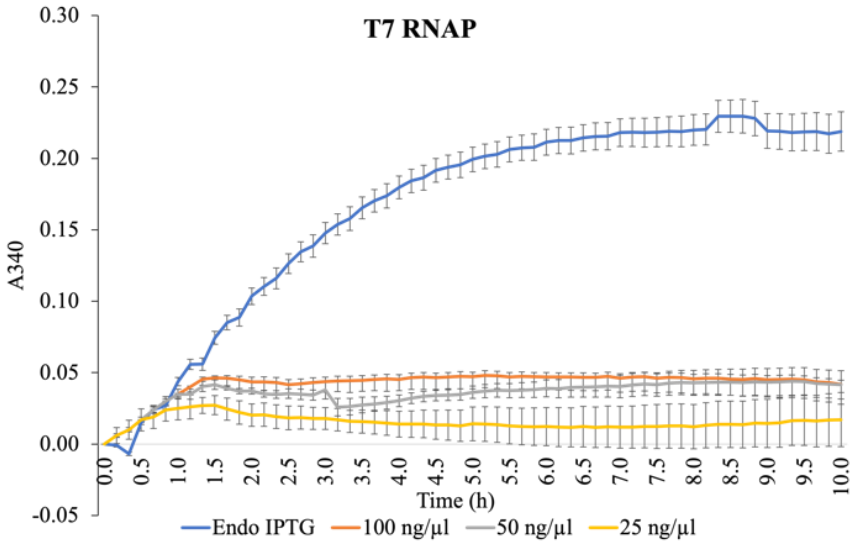
Reactions were initiated with increasing concentrations of inducers to determine the optimal concentration for the corresponding promoters. All reactions were prepared with BL21 Star(DE3) extracts that lack endogenous IPTG. (A, C) Increasing concentration of IPTG for pT7 promoter using plasmid DNA of optimized-*rppA* driven by pT7 promoter with the addition of 50 ng/µL T7 RNA polymerase. (B, D) Reactions were initiated with increasing concentrations of the extender unit malonyl-CoA concentration to determine optimal concentration using plasmid DNA of optimized-*rppA* driven by pT7, 50 ng/µL T7 RNA polymerase, and 500 µM IPTG. Reactions were run in triplicate and read every 10 mins for 6 hr. Error bars represent the standard error of the mean (n = 3). (C, D) Bar graph of (A, B) respectively.



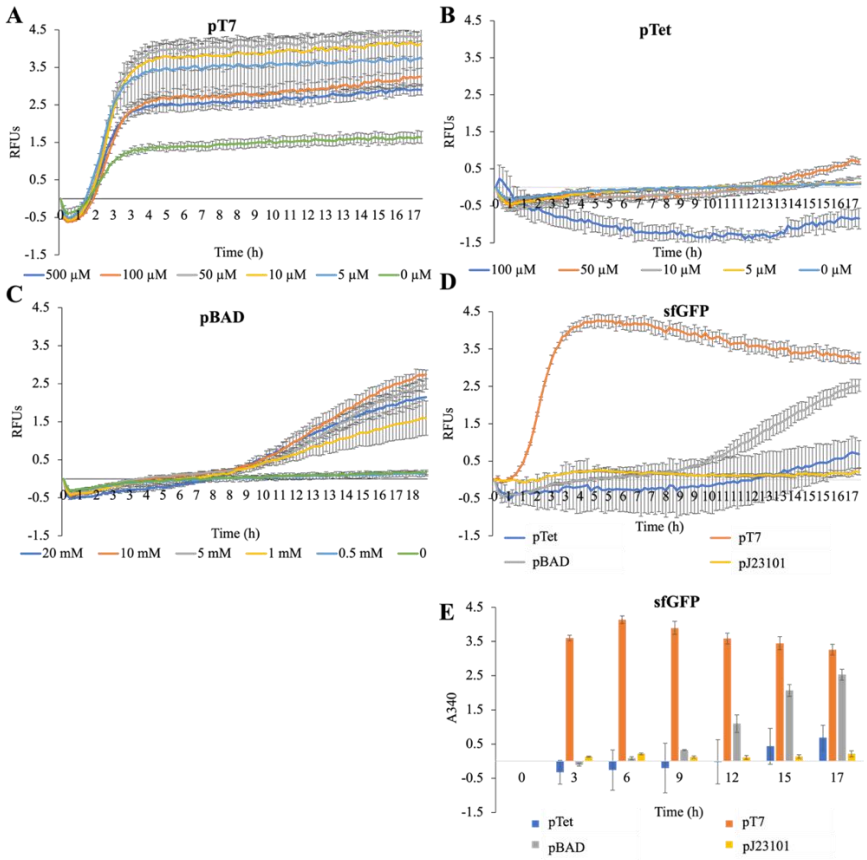
Appendix Figure S4.4. Optimization of anhydrotetracycline concentrations inducer for pTet promoter in *E. coli* lysate-based CFE reactions for RppA.
Reactions were initiated with increasing concentrations of anhydrotetracycline to determine the optimal concentration for pTet promoter. All reactions were prepared with BL21 Star(DE3) extracts that lack endogenous IPTG, plasmid DNA of optimized-*rppA* driven by pTet promoter, and 500 µM malonyl-CoA. Reactions were run in triplicate and read every 10 mins for 20 hr. Error bars represent the standard error of the mean (n = 3). (B) Bar graph of (A).



Appendix Figure S4.5. Optimization of L-arabinose concentrations inducer for pBAD promoter in *E. coli* lysate-based CFE reactions for RppA.
Reactions were initiated with increasing concentrations of L-arabinose to determine the optimal concentration for pBAD promoter. All reactions were prepared with BL21 Star(DE3) extracts that lack endogenous IPTG, plasmid DNA of optimized-*rppA* driven by pBAD promoter, and 500 µM malonyl-CoA. Reactions were run in triplicate and read every 10 mins for 20 hr. Error bars represent the standard error of the mean (n = 3). (B) Bar graph of (A).

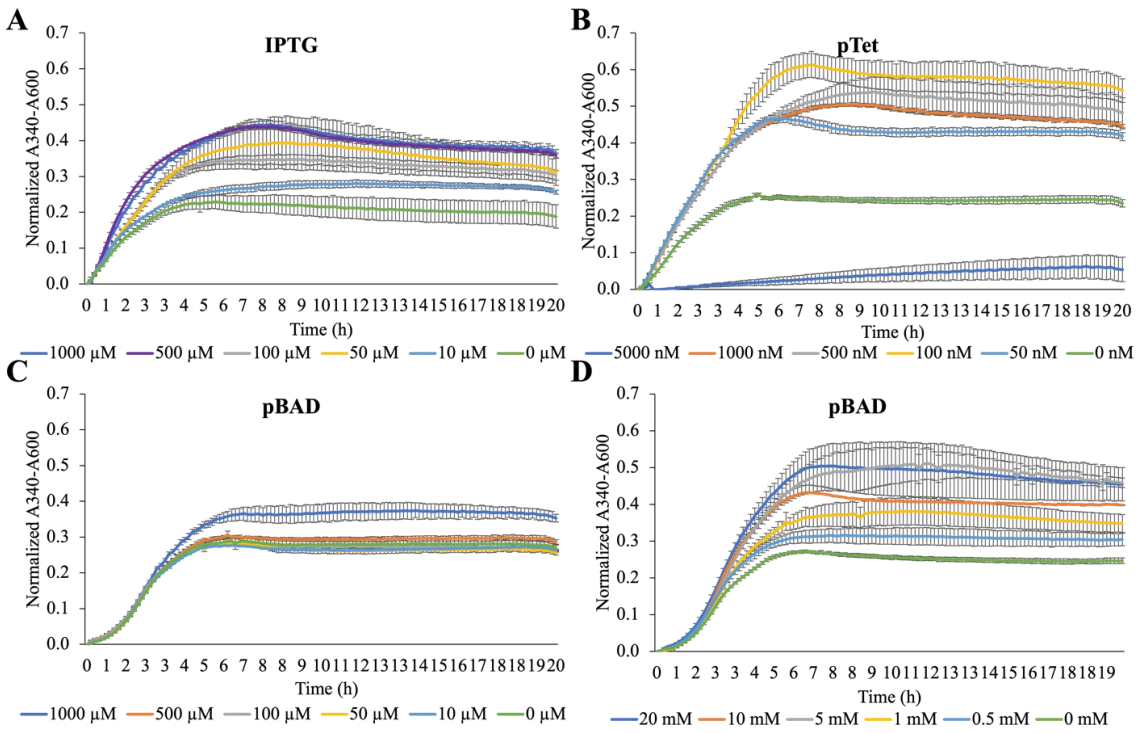


Appendix Figure S4.6. RppA production at different conditions of T7 RNA polymerase supplied.
Experiments were run by using pT7 constructs with optimized-*rppA* coding sequence for 1) lysate containing endogenous IPTG, 2) lysate does not contain endogenous IPTG, 500 µM IPTG, with increasing concentration of T7 RNA polymerase. Reactions were run in 6 replicates and read every 10 mins for 10 hr. Error bars represent the standard error of the mean (n = 6).



Appendix Figure S4.7. Optimization of inducer concentrations in *E. coli* lysate-based CFE reactions for sfGFP.

Reactions were initiated with increasing concentrations of inducers to determine the optimal concentration for the corresponding promoters. All reactions were prepared with BL21 Star(DE3) extracts that do not contain endogenous IPTG. (A) Increasing concentration of IPTG for pT7 promoter using plasmid DNA of sfGFP driven by pT7 with the addition of 50 ng/µL T7 RNA polymerase. (B) Increasing concentration of anhydrotetracycline for pTet promoter using plasmid DNA of sfGFP driven by pTet promoter. (C) Increasing concentration of L-arabinose for pBAD promoter using plasmid DNA of sfGFP driven by pBAD promoter. (D) CFE reaction of sfGFP in BL21 Star(DE3) extracts that lack endogenous IPTG. pT7 construct supplemented with 50 ng/µL T7 RNA polymerase and 500 µM IPTG. pTet construct supplemented with 50 µM anhydrotetracycline. pBAD construct supplemented with 10 mM L-arabinose. There is a delay in the expression of pTet and pBAD promoters, and both promoters indicate less expression than the pT7 promoter constructs. Reactions were run in triplicate and read every 10 mins for 20 hr. Error bars represent the standard error of the mean (n = 3). (E) Bar graph of (D).



Appendix Figure S4.8. Optimization of inducer concentrations in *E. coli in vivo* reactions.

Reactions were initiated with increasing concentrations of inducers to determine the optimal concentration for the corresponding promoters. All constructs were expressed in BL21 Star(DE3). (A) Reactions were initiated with increasing concentrations of IPTG using cells harboring plasmid DNA of HG-*rppA* driven by pT7. (B) Increasing concentration of tetracycline for pTet promoter using cells harboring plasmid DNA of HG-*rppA* driven by pTet. (C, D) Increasing concentration of L-arabinose for pBAD promoter using plasmid DNA of optimized-*rppA* driven by pBAD. Reactions were run in triplicate and read every 10 mins for 20 hr. Error bars represent the standard error of the mean (n = 3).

CHAPTER 5

CONCLUSIONS AND OUTLOOK

Parts of this chapter was originally published by Tien T Sword, Ghaeath S.K. Abbas, and Constance B Bailey:

Sword TT, Abbas GSK, Bailey CB. Cell-Free Protein Synthesis for Nonribosomal Peptide Synthetic Biology. Frontiers in Natural Products. 2024 Jan 18;

Natural products and their structural analogs have been and will continue to be the vital source of drugs, especially for the treatment of infectious diseases, and cancer. (252) However, screening and extracting natural products from their natural sources is often time-consuming, and laborious; which caused several drawbacks in the drug discovery industry in the 1900s. (253) While actinomycetes have historically been the major source of bioactive natural products, the traditional "grind and find" approaches of bioactivity-based screens and fractionation methods are found to be challenging. Evaluating actinomycetes multimodule proteins is a large bottleneck in synthetic biology because of their complex structures and their giant size. For these reasons, much research effort has been invested in developing strategies to express multimodule proteins in a high-yield and cost-effective manner. This work provides techniques to express multimodule proteins, improve the yields, and gain knowledge to help inform on what systems we should select to profile, characterize, and ultimately produce engineered metabolites to improve protein folding and/or activity.

Part I of this dissertation consists of **Chapter 2**, which describes the effects of refactoring strategies on megasynthases for heterologous expression in *E. coli.* Part II of this thesis encompasses **Chapters 3** and **4**, where cell-free protein synthesis strategies were applied to profile, investigate, and optimize biosynthetic enzyme expression for PKS/NRPS proteins.

One of the vital bottlenecks that restrain the development of megasynthase expression is the lack of knowledge about factors related to heterologous expression. Megasynthases from actinomycetal origin often fail to yield full-length, well-folded, and/or active proteins for heterologous expression. Thus, before we can make choices on what systems we should select to characterize, profile, and ultimately produce engineered metabolites, it is important to have a better fundamental understanding of the tradeoffs between genetic tractability and consequences on protein folding. In **Chapter 2**, a head-to-head comparison of codon-optimized sequences versus a native sequence of an NPRS Sreptomycete origin heterologously expressed in *E. coli* was performed. Albeit there have been numerous discussions about how codon usage affects the folding and solubilities of proteins, it is still not entirely clear what the consequences of codon bias in heterologous expression are. This work revealed the effect of codon usage strategies (synonymous coding or complementation of rare tRNAs) on protein yield and metabolite titer. The results suggested that the heterologous expression of less soluble proteins can be improved with a more sophisticated approach such as codon harmonization and/or the addition of rare tRNAs. These insights provide the scientific basis to justify using more sophisticated methods to express proteins heterologously with a substantial distance in CG content; and will lead to many future in-depth and detailed research directions such as developing a library of coding sequences for high GC content proteins.

Part II of this dissertation consists of **Chapters 3** and **4,** where the cell-free system was used as a tool to express multimodule enzyme Actinomycetele origin. Cell-free synthetic biology is a powerful platform that offers rapid, high-throughput, cost-effective, and robust approaches to access and harness many valuable natural products as well as generate analogs and apply the biosynthetic steps for bioproduct generation. CFPS accelerates the design-build-test-learn cycle of the metabolic engineering paradigm towards a full or semi-automated workflow from artisanal labor. This work focuses on using cell-free strategies to express PKS/NRPS enzymes, which opens the potential of using CFPS to accelerate the engineering of polyketide, non-ribosomal peptides, and their analogs.

Taking advantage of the flexibility of cell-free systems, exogenous elements can be rapidly added to the reaction directly and manipulated, allowing the synthesis of potential peptide natural products. **Chapter 3** details the investigation and optimization of a reporter-based assay in lysate-based CFE reactions to quickly sample different condition sets. Overall, this work reveals the relationship between enzyme structure and its catalytic activity versus CFE condition requirements extracts. Due to the large size, complex multidomain architecture, posttranslational modification, and cofactor requirements (e.g., ATP), NRPSs pose more demands to a cell-free system than simple fluorescent protein reporters. Thus, it is vital to tailor CFE reaction condition sets for different enzyme targets, especially for enzymes that interface with lysate endogenous metabolism. Using the tetracysteine tag enables quick measurement of full-length expressed proteins under different reaction conditions. This technique was also applied to optimize lysate-based conditions for NRPS proteins, contributing more tools to screen different engineered variants of novel heterologous pathways. With the increasing tools available to control, modify, and optimize conditions for CFPS systems with various lysate hosts, the promise of CFPS as a fruitful platform for bioengineering efforts remains high. For instance, the unique characteristics of the flexible linkers holding NRPS domains and modules together draw significant interest in studying these NRPS parts such as module and domain swapping, module and domain deletions, and insertions (37,254,255). Different approaches have been developed to engineer the A domain (37); however, expressing mutants often leads to low production or no production at all (254). Binding domains or modules are substantially longer than subdomains, hence, subdomain in swapping could greatly contribute to NRPS engineering based on gene synthesis and bioinformatics research. The subdomain swapping strategy was successfully done on the A domain of the GrsA biosynthesis of gramicidin S synthetase (37). Following, the production of the novel pyoverdine derivatives proved to be transferable to different NRPS systems at high yields by exchanging the A domain (150)(256). Another noticeable example of domain swapping is the exchanging of the T domain of the NRPS IndC with synthetic and natural T domains shows improvement in indigoidine production (153). Above are a few examples of domain engineering strategies, illuminating a promising approach to generating novel molecules as well as increasing natural product production.

Furthermore, identifying novel natural products from genome mining libraries and mutant natural products is a critical challenge due to the difficulty of expressing a huge number of gene clusters with high yields. Thus, establishing a robust and high-throughput method for the rapid expression in the screening process for large multienzyme is necessary. Albeit *E. coli* lysate-based CFPS systems have been developed for over twenty years to express various proteins with enhanced yields, (65) different proteins seem to have their own preferable expression conditions. (79) Besides optimizing conditions for existing commonly used DNA constructs, expanding plasmid DNA constructs with various refactoring strategies (e.g., DNA coding sequences, and promoter choices) can open more options to improve issues with complex protein expression such as solubility issues, codon usage bias, and post-translational modifications.

Chapter 4 leverages insights gained from **Chapter 2** and the method established in **Chapter 3** to screen a library of 16 plasmid DNA constructs of a type III PKS enzyme to examine expression strategies in a lysate-based cell-free system. Besides pT7 promoter systems and a selection of constitutive promoters, (248) other promoter systems and tools for codon harmonization have not been used extensively in CFE. This work explored the synergy effects between coding sequence versus promoter choices in CFE conditions. Broadly, refactoring promoters and/or coding sequences via CFE can be a valuable strategy to rapidly screen for catalytically functional production of enzymes from biosynthetic gene clusters. With this explosion of synthetic biology strategies, we indeed have an extensive toolkit to continue to mine the metabolic giftedness of actinomycetes for the discovery and development of new drug candidates. This can in turn accelerate DBTL cycles to generate valuable metabolites.

Developing strategies that improve the production of multimodular proteins will favor the drug discovery industry. While the scope of this dissertation in particular focuses on proteins of actinomycetal origin, other bacteria species harbor modular proteins with high GC content can be beneficial from tools and platform that was developed and investigated from this work. In summary, we anticipate the utility of CFPS platforms in the drug industry will be greatly expanded in the coming years. The flexibility of cell-free systems due to its open nature can greatly contribute to drug discovery and the creating of new analogs for the existing natural products.

REFERENCES

1. Harvey AL, Edrada-Ebel R, Quinn RJ. The re-emergence of natural products for drug discovery in the genomics era. Nat Rev Drug Discov. 2015 Feb;14(2):111–29.

2. Panda S, Zhou K. Engineering microbes to overproduce natural products as agrochemicals. Synthetic and Systems Biotechnology. 2023 Mar;8(1):79–85.

3. Sparks TC, Duke SO. Structure simplification of natural products as a lead generation approach in agrochemical discovery. J Agric Food Chem. 2021 Aug 4;69(30):8324–46.

4. Loiseleur O. Natural products in the discovery of agrochemicals. Chimia (Aarau). 2017 Dec 1;71(12):810–22.

5. Cantrell CL, Dayan FE, Duke SO. Natural products as sources for new pesticides. J Nat Prod. 2012 Jun 22;75(6):1231–42.

6. Romero EO, Saucedo AT, Hernández-Meléndez JR, Yang D, Chakrabarty S, Narayan ARH. Enabling broader adoption of biocatalysis in organic chemistry. JACS Au. 2023 Aug 28;3(8):2073–85.

7. Stout CN, Wasfy NM, Chen F, Renata H. Charting the evolution of chemoenzymatic strategies in the syntheses of complex natural products. J Am Chem Soc. 2023 Aug 23;145(33):18161–81.

8. Kries H, Trottmann F, Hertweck C. Novel Biocatalysts from Specialized Metabolism. Angew Chem Int Ed. 2023 Sep 22;e202309284.

9. Yuzawa S, Keasling JD, Katz L. Bio-based production of fuels and industrial chemicals by repurposing antibiotic-producing type I modular polyketide synthases: opportunities and challenges. J Antibiot. 2017 Apr;70(4):378–85.

10. Li X, Gadar-Lopez AE, Chen L, Jayachandran S, Cruz-Morales P, Keasling JD. Mining natural products for advanced biofuels and sustainable bioproducts. Curr Opin Biotechnol. 2023 Dec;84:103003.

11. Barajas JF, Blake-Hedges JM, Bailey CB, Curran S, Keasling JD. Engineered polyketides: Synergy between protein and host level engineering. Synthetic and Systems Biotechnology. 2017 Sep 7;2(3):147–66.

12. Atanasov AG, Waltenberger B, Pferschy-Wenzig E-M, Linder T, Wawrosch C, Uhrin P, et al. Discovery and resupply of pharmacologically active plant-derived natural products: A review. Biotechnol Adv. 2015 Dec;33(8):1582–614.

13. Clardy J, Fischbach MA, Currie CR. The natural history of antibiotics. Curr Biol. 2009 Jun 9;19(11):R437-41.

14. Ventola CL. The antibiotic resistance crisis: part 1: causes and threats. P T. 2015 Apr;40(4):277–83.

15. Newman DJ, Cragg GM. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. J Nat Prod. 2020 Mar 27;83(3):770–803.

16. Newman DJ, Cragg GM. Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod. 2012 Mar 23;75(3):311–35.

17. Newman DJ, Cragg GM. Natural Products as Sources of New Drugs from 1981 to 2014. J Nat Prod. 2016 Mar 25;79(3):629–61.

18. Patridge E, Gareiss P, Kinch MS, Hoyer D. An analysis of FDA-approved drugs: natural products and their derivatives. Drug Discov Today. 2016 Feb;21(2):204–7.

19. Baltz RH. Natural product drug discovery in the genomic era: realities, conjectures, misconceptions, and opportunities. J Ind Microbiol Biotechnol. 2019 Mar;46(3–4):281–99.

20. Wenski SL, Thiengmag S, Helfrich EJN. Complex peptide natural products: Biosynthetic principles, challenges and opportunities for pathway engineering. Synthetic and Systems Biotechnology. 2022 Mar;7(1):631–47.

21. Süssmuth RD, Mainz A. Nonribosomal Peptide Synthesis-Principles and Prospects. Angew Chem Int Ed. 2017 Mar 27;56(14):3770–821.

22. Felnagle EA, Jackson EE, Chan YA, Podevels AM, Berti AD, McMahon MD, et al. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. Mol Pharm. 2008 Apr;5(2):191–211.

23. Khosla C, Gokhale RS, Jacobsen JR, Cane DE. Tolerance and specificity of polyketide synthases. Annu Rev Biochem. 1999;68:219–53.

24. Mootz HD, Schwarzer D, Marahiel MA. Ways of Assembling Complex Natural Products on Modular Nonribosomal Peptide Synthetases. ChemBioChem. 2002 Jun 3;

25. Staunton J, Weissman KJ. Polyketide biosynthesis: a millennium review. Nat Prod Rep. 2001 Aug;18(4):380–416.

26. Weng J-K, Noel JP. Structure-function analyses of plant type III polyketide synthases. Meth Enzymol. 2012;515:317–35.

27. Rath CM, Scaglione JB, Kittendorf JD, Sherman DH. NRPS/PKS hybrid enzymes and their natural products. Comprehensive natural products II. Elsevier; 2010. p. 453–92.

28. Richardson M, Khosla C. Structure, function, and engineering of bacterial aromatic polyketide synthases. Comprehensive natural products chemistry. Elsevier; 1999. p. 473–94.

29. Chan YA, Podevels AM, Kevany BM, Thomas MG. Biosynthesis of polyketide synthase extender units. Nat Prod Rep. 2009 Jan;26(1):90–114.

30. Miller BR, Gulick AM. Structural biology of nonribosomal peptide synthetases. Methods Mol Biol. 2016;1401:3–29.

31. Barajas JF, Phelan RM, Schaub AJ, Kliewer JT, Kelly PJ, Jackson DR, et al. Comprehensive structural and biochemical analysis of the terminal myxalamid reductase domain for the engineered production of primary alcohols. Chem Biol. 2015 Aug 20;22(8):1018–29.

32. Heinemann H, Zhang H, Cox RJ. Reductive Release from a Hybrid PKS-NRPS during the Biosynthesis of Pyrichalasin H. Chem Eur J. 2023 Nov 5;e202302590.

33. Koryakina I, Kasey C, McArthur JB, Lowell AN, Chemler JA, Li S, et al. Inversion of extender unit selectivity in the erythromycin polyketide synthase by acyltransferase domain engineering. ACS Chem Biol. 2017 Jan 20;12(1):114–23.

34. Drufva EE, Hix EG, Bailey CB. Site directed mutagenesis as a precision tool to enable synthetic biology with engineered modular polyketide synthases. Synthetic and Systems Biotechnology. 2020 Jun;5(2):62–80.

35. Sun H, Liu Z, Zhao H, Ang EL. Recent advances in combinatorial biosynthesis for drug discovery. Drug Des Devel Ther. 2015 Feb 12;9:823–33.

36. Zhang J, Yan Y-J, An J, Huang S-X, Wang X-J, Xiang W-S. Designed biosynthesis of 25-methyl and 25-ethyl ivermectin with enhanced insecticidal activity by domain swap of avermectin polyketide synthase. Microb Cell Fact. 2015 Sep 24;14:152.

37. Kries H, Niquille DL, Hilvert D. A subdomain swap strategy for reengineering nonribosomal peptides. Chem Biol. 2015 May 21;22(5):640–8.

38. Khosla C, Zawada RJ. Generation of polyketide libraries via combinatorial biosynthesis. Trends Biotechnol. 1996 Sep;14(9):335–41.

39. Zarins-Tutt JS, Barberi TT, Gao H, Mearns-Spragg A, Zhang L, Newman DJ, et al. Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. Nat Prod Rep. 2016 Jan;33(1):54–72.

40. Stewart EJ. Growing unculturable bacteria. J Bacteriol. 2012 Aug;194(16):4151–60.

41. Dias DA, Urban S, Roessner U. A historical overview of natural products in drug discovery. Metabolites. 2012 Apr 16;2(2):303–36.

42. Jose PA, Maharshi A, Jha B. Actinobacteria in natural products research: Progress and prospects. Microbiol Res. 2021 May;246:126708.

43. Belknap KC, Park CJ, Barth BM, Andam CP. Genome mining of biosynthetic and chemotherapeutic gene clusters in Streptomyces bacteria. Sci Rep. 2020 Feb 6;10(1):2003.

44. Arai T, Takahashi K, Ishiguro K, Mikami Y. Some chemotherapeutic properties of two new antitumor antibiotics, saframycins A and C. Gan. 1980 Dec;71(6):790–6.

45. Martínez-Castro M, Barreiro C, Romero F, Fernández-Chimeno RI, Martín JF. Streptomyces tacrolimicus sp. nov., a low producer of the immunosuppressant tacrolimus (FK506). Int J Syst Evol Microbiol. 2011 May;61(Pt 5):1084–8.

46. Choi S-S, Nah H-J, Pyeon H-R, Kim E-S. Biosynthesis, regulation, and engineering of a linear polyketide tautomycetin: a novel immunosuppressant in Streptomyces sp. CK4412. J Ind Microbiol Biotechnol. 2017 May;44(4–5):555–61.

47. Kim HS, Park YI. Isolation and identification of a novel microorganism producing the immunosuppressant tacrolimus. J Biosci Bioeng. 2008 Apr;105(4):418–21.

48. Muramatsu H, Nagai K. Streptomyces tsukubensis sp. nov., a producer of the immunosuppressant tacrolimus. J Antibiot. 2013 Apr;66(4):251–4.

49. Castillo UF, Strobel GA, Ford EJ, Hess WM, Porter H, Jensen JB, et al. Munumbicins, wide-spectrum antibiotics produced by Streptomyces NRRL 30562, endophytic on Kennedia nigriscans. Microbiology (Reading, Engl). 2002 Sep;148(Pt 9):2675–85.

50. Production of an Antibiotic-like Activity by Streptomyces sp. COUK1 under Different Growth Conditions - ProQuest [Internet]. [cited 2023 Dec 12]. Available from: https://www.proquest.com/docview/1614204912?pq-origsite=gscholar&fromopenview=true

51. Thaker MN, Wang W, Spanogiannopoulos P, Waglechner N, King AM, Medina R, et al. Identifying producers of antibacterial compounds by screening for antibiotic resistance. Nat Biotechnol. 2013 Oct;31(10):922–7.

52. Vasilchenko AS, Julian WT, Lapchinskaya OA, Katrukha GS, Sadykova VS, Rogozhin EA. A Novel Peptide Antibiotic Produced by Streptomyces roseoflavus Strain INA-Ac-5812 With Directed Activity Against Gram-Positive Bacteria. Front Microbiol. 2020 Sep 15;11:556063.

53. Melbourne JK, Thompson KR, Peng H, Nixon K. Its complicated: The relationship between alcohol and microglia in the search for novel pharmacotherapeutic targets for alcohol use disorders. Prog Mol Biol Transl Sci. 2019 Jul 29;167:179–221.

54. Khajezadeh M, Abbaszadeh-Goudarzi K, Pourghadamyari H, Kafilzadeh F. A newly isolated Streptomyces rimosus strain capable of degrading deltamethrin as a pesticide in agricultural soil. J Basic Microbiol. 2020 May;60(5):435–43.

55. Bubici G. *Streptomyces* spp. as biocontrol agents against *Fusarium* species. CABI Reviews. 2018 Dec 5;1–15.

56. Abdel-Razek AS, El-Naggar ME, Allam A, Morsy OM, Othman SI. Microbial natural products in drug discovery. Processes. 2020 Apr 16;8(4):470.

57. Grohmann E, Muth G, Espinosa M. Conjugative plasmid transfer in gram-positive bacteria. Microbiol Mol Biol Rev. 2003 Jun;67(2):277–301, table of contents.

58. Musiol-Kroll EM, Tocchetti A, Sosio M, Stegmann E. Challenges and advances in genetic manipulation of filamentous actinomycetes - the remarkable producers of specialized metabolites. Nat Prod Rep. 2019 Sep 1;36(9):1351–69.

59. Li L, Maclntyre LW, Brady SF. Refactoring biosynthetic gene clusters for heterologous production of microbial natural products. Curr Opin Biotechnol. 2021 Jun;69:145–52.

60. Lawson CE, Harcombe WR, Hatzenpichler R, Lindemann SR, Löffler FE, O'Malley MA, et al. Common principles and best practices for engineering microbiomes. Nat Rev Microbiol. 2019 Dec;17(12):725–41.

61. Ji X, Liu W-Q, Li J. Recent advances in applying cell-free systems for high-value and complex natural product biosynthesis. Curr Opin Microbiol. 2022 Jun;67:102142.

62. Li J, Zhang L, Liu W. Cell-free synthetic biology for in vitro biosynthesis of pharmaceutical natural products. Synthetic and Systems Biotechnology. 2018 Jun;3(2):83–9.

63. Farag S, Bleich RM, Shank EA, Isayev O, Bowers AA, Tropsha A. Inter-Modular Linkers play a crucial role in governing the biosynthesis of non-ribosomal peptides. Bioinformatics. 2019 Oct 1;35(19):3584–91.

64. Zhou M, Guo J, Cha J, Chae M, Chen S, Barral JM, et al. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature. 2013 Mar 7;495(7439):111–5.

65. Carlson ED, Gan R, Hodgman CE, Jewett MC. Cell-free protein synthesis: applications come of age. Biotechnol Adv. 2012 Oct;30(5):1185–94.

66. Hodgman CE, Jewett MC. Cell-free synthetic biology: thinking outside the cell. Metab Eng. 2012 May;14(3):261–9.

67. Tuckey C, Asahara H, Zhou Y, Chong S. Protein synthesis using a reconstituted cell-free system. Curr Protoc Mol Biol. 2014 Oct 1;108:16.31.1-16.31.22.

68. Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, et al. Cell-free translation reconstituted with purified components. Nat Biotechnol. 2001 Aug;19(8):751–5.

69. Lavickova B, Maerkl SJ. A Simple, Robust, and Low-Cost Method To Produce the PURE Cell-Free System. ACS Synth Biol. 2019 Feb 15;8(2):455–62.

70. Cui Y, Chen X, Wang Z, Lu Y. Cell-Free PURE System: Evolution and Achievements. BioDesign Research. 2022 Aug 30;2022:9847014.

71. Wick S, Walsh DI, Bobrow J, Hamad-Schifferli K, Kong DS, Thorsen T, et al. PERSIA for Direct Fluorescence Measurements of Transcription, Translation, and Enzyme Activity in Cell-Free Systems. ACS Synth Biol. 2019 May 17;8(5):1010–25.

72. Wick S, Carr PA. Measurement of Transcription, Translation, and Other Enzymatic Processes During Cell-Free Expression Using PERSIA. Methods Mol Biol. 2022;2433:169–81.

73. Finking R, Solsbacher J, Konz D, Schobert M, Schafer A, Jahn D, et al. Characterization of a new type of phosphopantetheinyl transferase for fatty acid and siderophore synthesis in Pseudomonas aeruginosa. J Biol Chem. 2002 Dec 27;277(52):50293–302.

74. Bogart JW, Cabezas MD, Vögeli B, Wong DA, Karim AS, Jewett MC. Cell-Free Exploration of the Natural Product Chemical Space. Chembiochem. 2021 Jan 5;22(1):84–91.

75. Banerjee D, Eng T, Lau AK, Sasaki Y, Wang B, Chen Y, et al. Genome-scale metabolic rewiring improves titers rates and yields of the non-native product indigoidine at scale. Nat Commun. 2020 Oct 23;11(1):5385.

76. Gregorio NE, Levine MZ, Oza JP. A User's Guide to Cell-Free Protein Synthesis. Methods Protoc. 2019 Mar 12;2(1).

77. Goering AW, Li J, McClure RA, Thomson RJ, Jewett MC, Kelleher NL. In Vitro Reconstruction of Nonribosomal Peptide Biosynthesis Directly from DNA Using Cell-Free Protein Synthesis. ACS Synth Biol. 2017 Jan 20;6(1):39–44.

78. Zhuang L, Huang S, Liu W-Q, Karim AS, Jewett MC, Li J. Total in vitro biosynthesis of the nonribosomal macrolactone peptide valinomycin. Metab Eng. 2020 Jul;60:37–44.

79. Dinglasan JLN, Sword TT, Barker JW, Doktycz MJ, Bailey CB. Investigating and Optimizing the Lysate-Based Expression of Nonribosomal Peptide Synthetases Using a Reporter System. ACS Synth Biol. 2023 May 19;12(5):1447–60.

80. Moore SJ, Lai H-E, Chee S-M, Toh M, Coode S, Chengan K, et al. A Streptomyces venezuelae Cell-Free Toolkit for Synthetic Biology. ACS Synth Biol. 2021 Feb 19;10(2):402–11.

81. Siebels I, Nowak S, Heil CS, Tufar P, Cortina NS, Bode HB, et al. Cell-Free Synthesis of Natural Compounds from Genomic DNA of Biosynthetic Gene Clusters. ACS Synth Biol. 2020 Sep 18;9(9):2418–26.

82. Bundy BC, Hunt JP, Jewett MC, Swartz JR, Wood DW, Frey DD, et al. Cell-free biomanufacturing. Curr Opin Chem Eng. 2018 Dec;22:177–83.

83. Sezonov G, Joseleau-Petit D, D'Ari R. *Escherichia coli* physiology in Luria-Bertani broth. J Bacteriol. 2007 Dec;189(23):8746–9.

84. Pope B, Kent HM. High efficiency 5 min transformation of Escherichia coli. Nucleic Acids Res. 1996 Feb 1;24(3):536–7.

85. Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. Trends Biotechnol. 2004 Jul;22(7):346–53.

86. Lammertyn E, Van Mellaert L, Bijnens AP, Joris B, Anné J. Codon adjustment to maximise heterologous gene expression in Streptomyces lividans can lead to decreased mRNA stability and protein yield. Mol Gen Genet. 1996 Feb 5;250(2):223–9.

87. Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O. Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. Protein Expr Purif. 2008 May;59(1):94–102.

88. Seidle HF, Couch RD, Parry RJ. Characterization of a nonspecific phosphopantetheinyl transferase from Pseudomonas syringae pv. syringae FF5. Arch Biochem Biophys. 2006 Feb 15;446(2):167–74.

89. Garenne D, Haines MC, Romantseva EF, Freemont P, Strychalski EA, Noireaux V. Cell-free gene expression. Nat Rev Methods Primers. 2021 Dec;1(1):49.

90. Martin RW, Des Soye BJ, Kwon Y-C, Kay J, Davis RG, Thomas PM, et al. Cell-free protein synthesis from genomically recoded bacteria enables multisite

incorporation of noncanonical amino acids. Nat Commun. 2018 Mar 23;9(1):1203.

91. Jewett MC, Swartz JR. Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnol Bioeng. 2004 Apr 5;86(1):19–26.

92. Vilkhovoy M, Horvath N, Shih C-H, Wayman JA, Calhoun K, Swartz J, et al. Sequence Specific Modeling of E. coli Cell-Free Protein Synthesis. ACS Synth Biol. 2018 Aug 17;7(8):1844–57.

93. Hurst GB, Asano KG, Doktycz CJ, Consoli EJ, Doktycz WL, Foster CM, et al. Proteomics-Based Tools for Evaluation of Cell-Free Protein Synthesis. Anal Chem. 2017 Nov 7;89(21):11443–51.

94. Sword TT, Barker JW, Spradley M, Chen Y, Petzold CJ, Bailey CB. Expression of blue pigment synthetase a from Streptomyces lavenduale reveals insights on the effects of refactoring biosynthetic megasynthases for heterologous expression in Escherichia coli. Protein Expr Purif. 2023 Oct;210:106317.

95. Sword TT, Dinglasan JLN, Abbas GSK, William Barker J, Spradley ME, Greene ER, et al. Profiling Expression Strategies for a Type III Polyketide Synthase in a Lysate-Based, Cell-free System. BioRxiv. 2023 Dec 1;

96. Li JW-H, Vederas JC. Drug discovery and natural products: end of an era or an endless frontier? Science. 2009 Jul 10;325(5937):161–5.

97. Baltz RH. Gifted microbes for genome mining and natural product discovery. J Ind Microbiol Biotechnol. 2017 May;44(4–5):573–88.

98. Robbins T, Liu Y-C, Cane DE, Khosla C. Structure and mechanism of assembly line polyketide synthases. Curr Opin Struct Biol. 2016 Dec;41:10–8.

99. Hertweck C. Decoding and reprogramming complex polyketide assembly lines: prospects for synthetic biology. Trends Biochem Sci. 2015 Apr;40(4):189–99.

100. Cook TB, Pfleger BF. Leveraging synthetic biology for producing bioactive polyketides and non-ribosomal peptides in bacterial heterologous hosts. Medchemcomm. 2019 May 1;10(5):668–81.

101. Stevens DC, Hari TPA, Boddy CN. The role of transcription in heterologous expression of polyketides in bacterial hosts. Nat Prod Rep. 2013 Oct 11;30(11):1391–411.

102. Olano C, Lombó F, Méndez C, Salas JA. Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. Metab Eng. 2008 Sep;10(5):281–92.

103. Murli S, Kennedy J, Dayem LC, Carney JR, Kealey JT. Metabolic engineering of Escherichia coli for improved 6-deoxyerythronolide B production. J Ind Microbiol Biotechnol. 2003 Aug;30(8):500–9.

104. Palazzotto E, Tong Y, Lee SY, Weber T. Synthetic biology and metabolic engineering of actinomycetes for natural product discovery. Biotechnol Adv. 2019 Nov 1;37(6):107366.

105. Pickens LB, Tang Y, Chooi Y-H. Metabolic engineering for the production of natural products. Annu Rev Chem Biomol Eng. 2011;2:211–36.

106. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. Science. 2001 Mar 2;291(5509):1790–2.

107. Ikeda H, Kazuo S, Omura S. Genome mining of the Streptomyces avermitilis genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. J Ind Microbiol Biotechnol. 2014 Feb;41(2):233–50.

108. Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H. Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. Proc Natl Acad Sci USA. 2010 Feb 9;107(6):2646–51.

109. Li L, Liu X, Jiang W, Lu Y. Recent advances in synthetic biology approaches to optimize production of bioactive natural products in actinobacteria. Front Microbiol. 2019 Nov 5;10:2467.

110. Yuzawa S, Zargar A, Pang B, Katz L, Keasling JD. Commodity chemicals from engineered modular type I polyketide synthases. Meth Enzymol. 2018 May 26;608:393–415.

111. Pang B, Valencia LE, Wang J, Wan Y, Lal R, Zargar A, et al. Technical Advances to Accelerate Modular Type I Polyketide Synthase Engineering towards

a Retro-biosynthetic Platform. Biotechnol Bioprocess Eng. 2019 Jun;24(3):413–23.

112. Hopwood DA, Kieser T, Bibb M, Chater K. Practical Streptomyces Genetics | NHBS Academic & Professional Books [Internet]. 2000 [cited 2020 Nov 17]. Available from: https://www.nhbs.com/practical-streptomyces-genetics-book

113. Baltz RH. Genetic manipulation of secondary metabolite biosynthesis for improved production in Streptomyces and other actinomycetes. J Ind Microbiol Biotechnol. 2016 Mar;43(2–3):343–70.

114. Zhao Y, Li G, Chen Y, Lu Y. Challenges and advances in genome editing technologies in streptomyces. Biomolecules. 2020 May 8;10(5).

115. Lee N, Hwang S, Lee Y, Cho S, Palsson B, Cho B-K. Synthetic biology tools for novel secondary metabolite discovery in streptomyces. J Microbiol Biotechnol. 2019 May 28;29(5):667–86.

116. Myronovskyi M, Luzhetskyy A. Heterologous production of small molecules in the optimized Streptomyces hosts. Nat Prod Rep. 2019 Sep 1;36(9):1281–94.

117. Angov E, Legler PM, Mease RM. Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding. Methods Mol Biol. 2011;705:1–13.

118. Pellizza L, Smal C, Rodrigo G, Arán M. Codon usage clusters correlation: towards protein solubility prediction in heterologous expression systems in E. coli. Sci Rep. 2018 Jul 13;8(1):10618.

119. Angov E. Codon usage: nature's roadmap to expression and folding of proteins. Biotechnol J. 2011 Jun;6(6):650–9.

120. Zhou Z, Dang Y, Zhou M, Li L, Yu C-H, Fu J, et al. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. Proc Natl Acad Sci USA. 2016 Oct 11;113(41):E6117–25.

121. Wang Y, Li C, Khan MRI, Wang Y, Ruan Y, Zhao B, et al. An engineered rare codon device for optimization of metabolic pathways. Sci Rep. 2016 Feb 8;6:20608.

122. Edison LK, Dan VM, S. R R, N. S P. A Strategic Production Improvement of Streptomyces Beta Glucanase Enzymes with Aid of Codon Optimization and

Heterologous Expression. Biosci, Biotechnol Res Asia. 2020 Sep 20;17(03):587–99.

123. Takahashi H, Kumagai T, Kitani K, Mori M, Matoba Y, Sugiyama M. Cloning and characterization of a *Streptomyces* single module type non-ribosomal peptide synthetase catalyzing a blue pigment synthesis. J Biol Chem. 2007 Mar 23;282(12):9073–81.

124. Weissman KJ. The structural biology of biosynthetic megaenzymes. Nat Chem Biol. 2015 Sep;11(9):660–70.

125. Vickery CR, McCulloch IP, Sonnenschein EC, Beld J, Noel JP, Burkart MD. Dissecting modular synthases through inhibition: A complementary chemical and genetic approach. Bioorg Med Chem Lett. 2020 Jan 15;30(2):126820.

126. Hillson NJ, Rosengarten RD, Keasling JD. j5 DNA assembly design automation software. ACS Synth Biol. 2012 Jan 20;1(1):14–21.

127. Myers JA, Curtis BS, Curtis WR. Improving accuracy of cell and chromophore concentration measurements using optical density. BMC Biophys. 2013 Apr 22;6(1):4.

128. Wehrs M, Gladden JM, Liu Y, Platz L, Prahl J-P, Moon J, et al. Sustainable bioproduction of the blue pigment indigoidine: Expanding the range of heterologous products in *R. toruloides* to include non-ribosomal peptides. Green Chem. 2019;21(12):3394–406.

129. Pang B, Chen Y, Gan F, Yan C, Jin L, Gin JW, et al. Investigation of Indigoidine Synthetase Reveals a Conserved Active-Site Base Residue of Nonribosomal Peptide Synthetase Oxidases. J Am Chem Soc. 2020 Jun 24;142(25):10931–5.

130. Lo M-C, Aulabaugh A, Jin G, Cowling R, Bard J, Malamas M, et al. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. Anal Biochem. 2004 Sep 1;332(1):153–9.

131. Crowell AMJ, Wall MJ, Doucette AA. Maximizing recovery of water-soluble proteins through acetone precipitation. Anal Chim Acta. 2013 Sep 24;796:48–54.

132. Chen Y, Gin J, J Petzold C. Discovery proteomic (DDA) LC-MS/MS data acquisition and analysis v2. 2021 May 7;

133. Huynh K, Partch CL. Analysis of protein stability and ligand interactions by thermal shift assay. Curr Protoc Protein Sci. 2015 Feb 2;79:28.9.1-28.9.14.

134. Pless O, Kowenz-Leutz E, Dittmar G, Leutz A. A differential proteome screening system for post-translational modification-dependent transcription factor interactions. Nat Protoc. 2011 Mar;6(3):359–64.

135. Dorrestein PC, Bumpus SB, Calderone CT, Garneau-Tsodikova S, Aron ZD, Straight PD, et al. Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry. Biochemistry. 2006 Oct 24;45(42):12756–66.

136. Curran SC, Hagen A, Poust S, Chan LJG, Garabedian BM, de Rond T, et al. Probing the Flexibility of an Iterative Modular Polyketide Synthase with Non-Native Substrates *in Vitro*. ACS Chem Biol. 2018 Aug 17;13(8):2261–8.

137. Schilling B, Rardin MJ, MacLean BX, Zawadzka AM, Frewen BE, Cusack MP, et al. Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. Mol Cell Proteomics. 2012 May;11(5):202–14.

138. Vervoort EB, van Ravestein A, van Peij NN, Heikoop JC, van Haastert PJ, Verheijden GF, et al. Optimizing heterologous expression in dictyostelium: importance of 5’ codon adaptation. Nucleic Acids Res. 2000 May 15;28(10):2069–74.

139. Collart MA, Weiss B. Ribosome pausing, a dangerous necessity for co-translational events. Nucleic Acids Res. 2020 Feb 20;48(3):1043–55.

140. Moreira MH, Barros GC, Requião RD, Rossetto S, Domitrovic T, Palhano FL. From reporters to endogenous genes: the impact of the first five codons on translation efficiency in Escherichia coli. RNA Biol. 2019 Dec;16(12):1806–16.

141. Skiba MA, Maloney FP, Dan Q, Fraley AE, Aldrich CC, Smith JL, et al. PKS-NRPS Enzymology and Structural Biology: Considerations in Protein Production. Meth Enzymol. 2018 Mar 16;604:45–88.

142. Oresic M, Shalloway D. Specific correlations between relative synonymous codon usage and protein secondary structure. J Mol Biol. 1998 Aug 7;281(1):31–48.

143. Mignon C, Mariano N, Stadthagen G, Lugari A, Lagoutte P, Donnat S, et al. Codon harmonization - going beyond the speed limit for protein expression. FEBS Lett. 2018 May;592(9):1554–64.

144. Angov E, Hillier CJ, Kincaid RL, Lyon JA. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. PLoS ONE. 2008 May 14;3(5):e2189.

145. Martínez-Núñez MA, López VEL y. Nonribosomal peptides synthetases and their applications in industry. Sustain Chem Process. 2016 Dec;4(1):13.

146. Wang H, Fewer DP, Holm L, Rouhiainen L, Sivonen K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. Proc Natl Acad Sci USA. 2014 Jun 24;111(25):9259–64.

147. Ishikawa F, Nohara M, Nakamura S, Nakanishi I, Tanabe G. Precise probing of residue roles by NRPS code swapping: mutation, enzymatic characterization, modeling, and substrate promiscuity of aryl acid adenylation domains. Biochemistry. 2020 Feb 4;59(4):351–63.

148. Bian X, Plaza A, Yan F, Zhang Y, Müller R. Rational and efficient site-directed mutagenesis of adenylation domain alters relative yields of luminmide derivatives in vivo. Biotechnol Bioeng. 2015 Jul;112(7):1343–53.

149. Stachelhaus T, Mootz HD, Marahiel MA. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem Biol. 1999 Aug;6(8):493–505.

150. Calcott MJ, Owen JG, Ackerley DF. Efficient rational modification of non-ribosomal peptides by adenylation domain substitution. Nat Commun. 2020 Sep 11;11(1):4554.

151. Bozhüyük KAJ, Fleischhacker F, Linck A, Wesche F, Tietze A, Niesert C-P, et al. De novo design and engineering of non-ribosomal peptide synthetases. Nat Chem. 2018 Mar;10(3):275–81.

152. Chiocchini C, Linne U, Stachelhaus T. In vivo biocombinatorial synthesis of lipopeptides by COM domain-mediated reprogramming of the surfactin biosynthetic complex. Chem Biol. 2006 Aug;13(8):899–908.

153. Beer R, Herbst K, Ignatiadis N, Kats I, Adlung L, Meyer H, et al. Creating functional engineered variants of the single-module non-ribosomal peptide synthetase IndC by T domain exchange. Mol Biosyst. 2014 Jul;10(7):1709–18.

154. Kries H. Biosynthetic engineering of nonribosomal peptide synthetases. J Pept Sci. 2016 Sep;22(9):564–70.

155. Kries H, Wachtel R, Pabst A, Wanner B, Niquille D, Hilvert D. Reprogramming nonribosomal peptide synthetases for "clickable" amino acids. Angew Chem Int Ed. 2014 Sep 15;53(38):10105–8.

156. Bozhüyük KAJ, Linck A, Tietze A, Kranz J, Wesche F, Nowak S, et al. Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. Nat Chem. 2019 Jul;11(7):653–61.

157. Thirlway J, Lewis R, Nunns L, Al Nakeeb M, Styles M, Struck A-W, et al. Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. Angew Chem Int Ed. 2012 Jul 16;51(29):7181–4.

158. Kudo F, Miyanaga A, Eguchi T. Structural basis of the nonribosomal codes for nonproteinogenic amino acid selective adenylation enzymes in the biosynthesis of natural products. J Ind Microbiol Biotechnol. 2019 Mar;46(3–4):515–36.

159. Villiers B, Hollfelder F. Directed evolution of a gatekeeper domain in nonribosomal peptide synthesis. Chem Biol. 2011 Oct 28;18(10):1290–9.

160. Challis GL, Naismith JH. Structural aspects of non-ribosomal peptide biosynthesis. Curr Opin Struct Biol. 2004 Dec;14(6):748–56.

161. Dopp JL, Jo YR, Reuel NF. Methods to reduce variability in E. Coli-based cell-free protein expression experiments. Synthetic and Systems Biotechnology. 2019 Dec;4(4):204–11.

162. Dopp BJL, Tamiev DD, Reuel NF. Cell-free supplement mixtures: Elucidating the history and biochemical utility of additives used to support in vitro protein synthesis in E. coli extract. Biotechnol Adv. 2019;37(1):246–58.

163. Li J, Wang H, Kwon Y-C, Jewett MC. Establishing a high yielding streptomyces-based cell-free protein synthesis system. Biotechnol Bioeng. 2017 Jun;114(6):1343–53.

164. Li J, Wang H, Jewett MC. Expanding the palette of Streptomyces -based cell-free protein synthesis systems with enhanced yields. Biochem Eng J. 2018 Feb;130:29–33.

165. Moore SJ, Lai H-E, Needham H, Polizzi KM, Freemont PS. Streptomyces venezuelae TX-TL - a next generation cell-free synthetic biology tool. Biotechnol J. 2017 Apr;12(4).

166. Kelwick R, Webb AJ, MacDonald JT, Freemont PS. Development of a Bacillus subtilis cell-free transcription-translation system for prototyping regulatory elements. Metab Eng. 2016 Nov;38:370–81.

167. Owen JG, Copp JN, Ackerley DF. Rapid and flexible biochemical assays for evaluating 4'-phosphopantetheinyl transferase activity. Biochem J. 2011 Jun 15;436(3):709–17.

168. Owen JG, Calcott MJ, Robins KJ, Ackerley DF. Generating functional recombinant NRPS enzymes in the laboratory setting via peptidyl carrier protein engineering. Cell Chem Biol. 2016 Nov 17;23(11):1395–406.

169. Hoffmann C, Gaietta G, Zürn A, Adams SR, Terrillon S, Ellisman MH, et al. Fluorescent labeling of tetracysteine-tagged proteins in intact cells. Nat Protoc. 2010 Sep 23;5(10):1666–77.

170. Stevens DC, Conway KR, Pearce N, Villegas-Peñaranda LR, Garza AG, Boddy CN. Alternative sigma factor over-expression enables heterologous expression of a type II polyketide biosynthetic pathway in Escherichia coli. PLoS ONE. 2013 May 28;8(5):e64858.

171. Brown AS, Robins KJ, Ackerley DF. A sensitive single-enzyme assay system using the non-ribosomal peptide synthetase BpsA for measurement of L-glutamine in biological samples. Sci Rep. 2017 Jan 31;7:41745.

172. Auld DS, Coassin PA, Coussens NP, Hensley P, Klumpp-Thomas C, Michael S, et al. Microplate Selection and Recommended Practices in High-throughput Screening and Quantitative Biology. In: Sittampalam GS, Grossman A, Brimacombe K, Arkin M, Auld D, Austin CP, et al., editors. Assay Guidance Manual. Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2004.

173. Liang Y, Woodle SA, Shibeko AM, Lee TK, Ovanesov MV. Correction of microplate location effects improves performance of the thrombin generation test. Thromb J. 2013 Jul 5;11(1):12.

174. Kim DM, Swartz JR. Prolonging cell-free protein synthesis with a novel ATP regeneration system. Biotechnol Bioeng. 1999;66(3):180–8.

175. Yang C, Yang M, Zhao W, Ding Y, Wang Y, Li J. Establishing a Klebsiella pneumoniae-Based Cell-Free Protein Synthesis System. Molecules. 2022 Jul 22;27(15).

176. Des Soye BJ, Davidson SR, Weinstock MT, Gibson DG, Jewett MC. Establishing a High-Yielding Cell-Free Protein Synthesis Platform Derived from Vibrio natriegens. ACS Synth Biol. 2018 Sep 21;7(9):2245–55.

177. Garcia DC, Dinglasan JLN, Shrestha H, Abraham PE, Hettich RL, Doktycz MJ. A lysate proteome engineering strategy for enhancing cell-free metabolite production. Metab Eng Commun. 2021 Jun;12:e00162.

178. Dinglasan JLN, Reeves DT, Hettich RL, Doktycz MJ. Liquid Chromatography Coupled to Refractive Index or Mass Spectrometric Detection for Metabolite Profiling in Lysate-based Cell-free Systems. J Vis Exp. 2021 Sep 23;(175).

179. Conti E, Stachelhaus T, Marahiel MA, Brick P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. EMBO J. 1997 Jul 16;16(14):4174–83.

180. Guo X, Zhu Y, Bai L, Yang D. The Protection Role of Magnesium Ions on Coupled Transcription and Translation in Lyophilized Cell-Free System. ACS Synth Biol. 2020 Apr 17;9(4):856–63.

181. Failmezger J, Scholz S, Blombach B, Siemann-Herzberg M. Cell-Free Protein Synthesis From Fast-Growing Vibrio natriegens. Front Microbiol. 2018 Jun 1;9:1146.

182. Kim J, Copeland CE, Seki K, Vögeli B, Kwon Y-C. Tuning the Cell-Free Protein Synthesis System for Biomanufacturing of Monomeric Human Filaggrin. Front Bioeng Biotechnol. 2020 Oct 29;8:590341.

183. Whittaker JW. Cell-free protein synthesis: the state of the art. Biotechnol Lett. 2013 Feb;35(2):143–52.

184. Kim DM, Swartz JR. Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. Biotechnol Bioeng. 2001 Aug 20;74(4):309–16.

185. Voloshin AM, Swartz JR. Large-Scale Batch Reactions for Cell-Free Protein Synthesis. In: Spirin AS, Swartz JR, editors. Cell-Free Protein Synthesis. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2007. p. 207–35.

186. Calhoun KA, Swartz JR. Total amino acid stabilization during cell-free protein synthesis reactions. J Biotechnol. 2006 May 17;123(2):193–203.

187. Hahn M, Stachelhaus T. Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. Proc Natl Acad Sci USA. 2004 Nov 2;101(44):15585–90.

188. Klapper M, Braga D, Lackner G, Herbst R, Stallforth P. Bacterial alkaloid biosynthesis: structural diversity via a minimalistic nonribosomal peptide synthetase. Cell Chem Biol. 2018 Jun 21;25(6):659-665.e9.

189. Gulick AM, Aldrich CC. Trapping interactions between catalytic domains and carrier proteins of modular biosynthetic enzymes with chemical probes. Nat Prod Rep. 2018 Nov 14;35(11):1156–84.

190. Bozhüyük KA, Micklefield J, Wilkinson B. Engineering enzymatic assembly lines to produce new antibiotics. Curr Opin Microbiol. 2019 Oct;51:88–96.

191. Bérdy J. Bioactive microbial metabolites. J Antibiot. 2005 Jan;58(1):1–26.

192. Drufva EE, Sword TT, Bailey CB. Metabolic engineering of actinomycetes for natural product discovery. In: Rai RV, Bai JA, editors. Natural Products from Actinomycetes: Diversity, Ecology and Drug Discovery. Singapore: Springer Singapore; 2022. p. 267–307.

193. Schmidt M, Lee N, Zhan C, Roberts JB, Nava AA, Keiser LS, et al. Maximizing heterologous expression of engineered type I polyketide synthases: investigating codon optimization strategies. ACS Synth Biol. 2023 Nov 17;12(11):3366–80.

194. Dubendorff JW, Studier FW. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. J Mol Biol. 1991 May 5;219(1):45–59.

195. William Studier F, Rosenberg AH, Dunn JJ, Dubendorff JW. [6] Use of T7 RNA polymerase to direct expression of cloned genes. Gene Expression Technology. Elsevier; 1990. p. 60–89.

196. Wagner L, Jules M, Borkowski O. What remains from living cells in bacterial lysate-based cell-free systems. Comput Struct Biotechnol J. 2023 May 24;21:3173–82.

197. Dinglasan JLN, Doktycz MJ. Rewiring Cell-free Metabolic Flux in *E. coli* Lysates using a Block-Push-Pull Approach. Synth Biol. 2023 Apr 17;

198. Mouncey NJ, Otani H, Udwary D, Yoshikuni Y. New voyages to explore the natural product galaxy. J Ind Microbiol Biotechnol. 2019 Mar;46(3–4):273–9.

199. Moore SJ, MacDonald JT, Wienecke S, Ishwarbhai A, Tsipa A, Aw R, et al. Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria. Proc Natl Acad Sci USA. 2018 May 8;115(19):E4340–9.

200. Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol. 2006 Jan;24(1):79–88.

201. Lentini R, Forlin M, Martini L, Del Bianco C, Spencer AC, Torino D, et al. Fluorescent proteins and in vitro genetic organization for cell-free synthetic biology. ACS Synth Biol. 2013 Sep 20;2(9):482–9.

202. Jew K, Smith PEJ, So B, Kasman J, Oza JP, Black MW. Characterizing and Improving pET Vectors for Cell-free Expression. Front Bioeng Biotechnol. 2022 Jun 23;10:895069.

203. Burrington LR, Watts KR, Oza JP. Characterizing and Improving Reaction Times for E. coli-Based Cell-Free Protein Synthesis. ACS Synth Biol. 2021 Aug 20;10(8):1821–9.

204. Funa N, Ohnishi Y, Fujii I, Shibuya M, Ebizuka Y, Horinouchi S. A new pathway for polyketide synthesis in microorganisms. Nature. 1999 Aug 26;400(6747):897–9.

205. Funa N, Ohnishi Y, Ebizuka Y, Horinouchi S. Properties and substrate specificity of RppA, a chalcone synthase-related polyketide synthase in Streptomyces griseus. J Biol Chem. 2002 Feb 15;277(7):4628–35.

206. Yang D, Kim WJ, Yoo SM, Choi JH, Ha SH, Lee MH, et al. Repurposing type III polyketide synthase as a malonyl-CoA biosensor for metabolic engineering in bacteria. Proc Natl Acad Sci USA. 2018 Oct 2;115(40):9835–44.

207. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol. 1995 Jul;177(14):4121–30.

208. Lee SK, Newman JD, Keasling JD. Catabolite repression of the propionate catabolic genes in Escherichia coli and Salmonella enterica: evidence for involvement of the cyclic AMP receptor protein. J Bacteriol. 2005 Apr;187(8):2793–800.

209. Lee TS, Krupa RA, Zhang F, Hajimorad M, Holtz WJ, Prasad N, et al. BglBrick vectors and datasheets: A synthetic biology platform for gene expression. J Biol Eng. 2011 Sep 20;5:12.

210. Krefft D, Papkov A, Zylicz-Stachula A, Skowron PM. Thermostable proteins bioprocesses: The activity of restriction endonuclease-methyltransferase from Thermus thermophilus (RM.TthHB27I) cloned in Escherichia coli is critically affected by the codon composition of the synthetic gene. PLoS ONE. 2017 Oct 17;12(10):e0186633.

211. Walsh IM, Bowman MA, Soto Santarriaga IF, Rodriguez A, Clark PL. Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. Proc Natl Acad Sci USA. 2020 Feb 18;117(7):3528–34.

212. Chaney JL, Steele A, Carmichael R, Rodriguez A, Specht AT, Ngo K, et al. Widespread position-specific conservation of synonymous rare codons within coding sequences. PLoS Comput Biol. 2017 May 5;13(5):e1005531.

213. Welch M, Govindarajan S, Ness JE, Villalobos A, Gurney A, Minshull J, et al. Design parameters to control synthetic gene expression in Escherichia coli. PLoS ONE. 2009 Sep 14;4(9):e7002.

214. Mellitzer A, Weis R, Glieder A, Flicker K. Expression of lignocellulolytic enzymes in Pichia pastoris. Microb Cell Fact. 2012 May 14;11:61.

215. Kodumal SJ, Patel KG, Reid R, Menzella HG, Welch M, Santi DV. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. Proc Natl Acad Sci USA. 2004 Nov 2;101(44):15573–8.

216. Feng Z, Zhang L, Han X, Zhang Y. Codon optimization of the calf prochymosin gene and its expression in Kluyveromyces lactis. World J Microbiol Biotechnol. 2010 May;26(5):895–901.

217. Marlatt NM, Spratt DE, Shaw GS. Codon optimization for enhanced Escherichia coli expression of human S100A11 and S100A1 proteins. Protein Expr Purif. 2010 Sep;73(1):58–64.

218. Villalobos A, Ness JE, Gustafsson C, Minshull J, Govindarajan S. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinformatics. 2006 Jun 6;7:285.

219. Richardson SM, Wheelan SJ, Yarrington RM, Boeke JD. GeneDesign: rapid, automated design of multikilobase synthetic genes. Genome Res. 2006 Apr;16(4):550–6.

220. Wright G, Rodriguez A, Li J, Milenkovic T, Emrich SJ, Clark PL. CHARMING: Harmonizing synonymous codon usage to replicate a desired codon usage pattern. Protein Sci. 2022 Jan;31(1):221–31.

221. Spencer PS, Siller E, Anderson JF, Barral JM. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. J Mol Biol. 2012 Sep 21;422(3):328–35.

222. Shabalina SA, Spiridonov NA, Kashina A. Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. Nucleic Acids Res. 2013 Feb 1;41(4):2073–94.

223. Gilchrist MA, Chen W-C, Shah P, Landerer CL, Zaretzki R. Estimating Gene Expression and Codon-Specific Translational Efficiencies, Mutation Biases, and Selection Coefficients from Genomic Data Alone. Genome Biol Evol. 2015 May 14;7(6):1559–79.

224. Clarke TF, Clark PL. Rare codons cluster. PLoS ONE. 2008 Oct 15;3(10):e3412.

225. Rodriguez A, Wright G, Emrich S, Clark PL. %MinMax: A versatile tool for calculating and comparing synonymous codon usage and its impact on protein folding. Protein Sci. 2018 Jan;27(1):356–62.

226. Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res. 2000 Jan 1;28(1):292.

227. Cope AL, Gilchrist MA. Quantifying shifts in natural selection on codon usage between protein regions: a population genetics approach. BMC Genomics. 2022 May 30;23(1):408.

228. McKevitt M, Patel K, Smajs D, Marsh M, McLoughlin M, Norris SJ, et al. Systematic cloning of Treponema pallidum open reading frames for protein expression and antigen discovery. Genome Res. 2003 Jul;13(7):1665–74.

229. Senda N, Enomoto T, Kihara K, Yamashiro N, Takagi N, Kiga D, et al. Development of an expression-tunable multiple protein synthesis system in cell-free reactions using T7-promoter-variant series. Synth Biol (Oxf). 2022 Nov 25;7(1):ysac029.

230. Karim AS, Liew FE, Garg S, Vögeli B, Rasor BJ, Gonnot A, et al. Modular cell-free expression plasmids to accelerate biological design in cells. Synth Biol (Oxf). 2020 Oct 14;5(1):ysaa019.

231. Swartz JR, Jewett MC, Woodrow KA. Cell-free protein synthesis with prokaryotic combined transcription-translation. Methods Mol Biol. 2004;267:169–82.

232. Sun ZZ, Hayes CA, Shin J, Caschera F, Murray RM, Noireaux V. Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. J Vis Exp. 2013 Sep 16;(79):e50762.

233. Garenne D, Thompson S, Brisson A, Khakimzhan A, Noireaux V. The all-E. coliTXTL toolbox 3.0: new capabilities of a cell-free synthetic biology platform. Synth Biol (Oxf). 2021 Aug 4;6(1):ysab017.

234. Tokmakov AA, Fukami Y. Activation of T7 RNA polymerase in Xenopus oocytes and cell-free extracts. Genes Cells. 2010 Nov;15(11):1136–44.

235. Hagen A, Poust S, de Rond T, Yuzawa S, Katz L, Adams PD, et al. In vitro analysis of carboxyacyl substrate tolerance in the loading and first extension modules of borrelidin polyketide synthase. Biochemistry. 2014 Sep 30;53(38):5975–7.

236. Hagen A, Poust S, Rond T de, Fortman JL, Katz L, Petzold CJ, et al. Engineering a polyketide synthase for in vitro production of adipic acid. ACS Synth Biol. 2016 Jan 15;5(1):21–7.

237. Karig DK, Iyer S, Simpson ML, Doktycz MJ. Expression optimization and synthetic gene networks in cell-free systems. Nucleic Acids Res. 2012 Apr;40(8):3763–74.

238. Borkowski O, Bricio C, Murgiano M, Rothschild-Mancinelli B, Stan G-B, Ellis T. Cell-free prediction of protein expression costs for growing cells. Nat Commun. 2018 Apr 13;9(1):1457.

239. Brooks R, Morici L, Sandoval N. Cell Free Bacteriophage Synthesis from Engineered Strains Improves Yield. ACS Synth Biol. 2023 Aug 18;12(8):2418–31.

240. Guo S, Murray RM. Construction of Incoherent Feedforward Loop Circuits in a Cell-Free System and in Cells. ACS Synth Biol. 2019 Mar 15;8(3):606–10.

241. Levine MZ, Gregorio NE, Jewett MC, Watts KR, Oza JP. Escherichia coli-Based Cell-Free Protein Synthesis: Protocols for a robust, flexible, and accessible platform technology. J Vis Exp. 2019 Feb 25;(144).

242. Takamura Y, Nomura G. Changes in the intracellular concentration of acetyl-CoA and malonyl-CoA in relation to the carbon and energy metabolism of Escherichia coli K12. J Gen Microbiol. 1988 Aug;134(8):2249–53.

243. Karim AS, Dudley QM, Juminaga A, Yuan Y, Crowe SA, Heggestad JT, et al. In vitro prototyping and rapid optimization of biosynthetic enzymes for cell design. Nat Chem Biol. 2020 Aug;16(8):912–9.

244. Vögeli B, Schulz L, Garg S, Tarasava K, Clomburg JM, Lee SH, et al. Cell-free prototyping enables implementation of optimized reverse β-oxidation pathways in heterotrophic and autotrophic bacteria. Nat Commun. 2022 Jun 1;13(1):3058.

245. Sivashanmugam A, Murray V, Cui C, Zhang Y, Wang J, Li Q. Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. Protein Sci. 2009 May;18(5):936–48.

246. Geurink PP, van der Heden van Noort GJ, Mulder MPC, Knaap RCM, Kikkert M, Ovaa H. Profiling DUBs and Ubl-specific proteases with activity-based probes. Meth Enzymol. 2019 Feb 14;618:357–87.

247. Khlebnikov A, Risa O, Skaug T, Carrier TA, Keasling JD. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. J Bacteriol. 2000 Dec;182(24):7029–34.

248. Chappell J, Jensen K, Freemont PS. Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. Nucleic Acids Res. 2013 Mar 1;41(5):3471–81.

249. Wright G, Rodriguez A, Clark PL, Emrich S. A new look at codon usage and protein expression. Epic Ser Comput. 2019 Mar 18;60:104–12.

250. Medina R, Lucentini CG, Franco MEE, Petroselli G, Rosso JA, Erra-Balsells R, et al. Identification of an intermediate for 1,8-dihydroxynaphthalene-melanin synthesis in a race-2 isolate of Fulvia fulva (syn. Cladosporium fulvum). Heliyon. 2018 Dec 17;4(12):e01036.

251. Becker K, Lambert C, Wieschhaus J, Stadler M. Phylogenetic Assignment of the Fungicolous Hypoxylon invadens (Ascomycota, Xylariales) and Investigation of its Secondary Metabolites. Microorganisms. 2020 Sep 11;8(9).

252. Newman DJ, Cragg GM. Natural products as sources of new drugs over the last 25 years. J Nat Prod. 2007 Mar;70(3):461–77.

253. Atanasov AG, Zotchev SB, Dirsch VM, International Natural Product Sciences Taskforce, Supuran CT. Natural products in drug discovery: advances and opportunities. Nat Rev Drug Discov. 2021 Mar;20(3):200–16.

254. Winn M, Fyans JK, Zhuo Y, Micklefield J. Recent advances in engineering nonribosomal peptide assembly lines. Nat Prod Rep. 2016 Feb;33(2):317–47.

255. Beck C, Garzón JFG, Weber T. Recent Advances in Re-engineering Modular PKS and NRPS Assembly Lines. Biotechnol Bioprocess Eng. 2020 Dec;25(6):886–94.

256. Messenger SR, McGuinniety EMR, Stevenson LJ, Owen JG, Challis GL, Ackerley DF, et al. Metagenomic domain substitution for the high-throughput modification of nonribosomal peptides. Nat Chem Biol. 2023 Nov 23;

VITA

Tien Thuy Tran was born and raised in Danang city, Vietnam, where she attended Phan Chau Trinh High School. After High School, she moved to the U.S to pursue her career in science. She received her Bachelor of Science in Chemistry with a minor in Biology from the University of Tennessee, Knoxville in 2019. As an undergraduate, she pursued a variety of research areas from different laboratories at the University of Tennessee including: Microbiology laboratory (under the mentorship of Prof. Terry C. Hazen) working on microorganism in water columns and sediments of the Caspian Sea, Organic Chemistry laboratory (under the mentorship of Prof. Michael D. Best) working on identifying and characterizing organic compounds, and Biochemistry laboratory (under the mentorship of Prof. Constance B. Bailey) working on different heterologous hosts expression strategies for megasynthase proteins. For graduate school, she continued her research in the field of natural product biosynthesis at the Bailey laboratory.

This dissertation was typed by Tien Thuy Tran.