

Profiling Expression Strategies for a Type III Polyketide Synthase in a Lysate-Based, Cell-free System

Authors: Tien T. Sword^{‡, a}, Jaime Lorenzo N. Dinglasan^{‡, b, c}, Ghaeath S.K. Abbas,^{a, d} J. William Barker,^a Madeline E. Spradley,^c Elijah R. Greene,^a Damian S. Gooden,^a Scott J. Emrich,^{c, f, g} Michael A. Gilchrist,^{c, g} Mitchel J. Doktycz,^{*b, c} Constance B. Bailey^{*a, c, d}

Author affiliations:

*Corresponding authors

Contact: doktyczmj@ornl.gov, Constance.Bailey@Sydney.edu.au

^a Department of Chemistry, University of Tennessee-Knoxville (Knoxville, TN USA)

^b Biosciences Division, Oak Ridge National Laboratory (Oak Ridge, TN USA)

^c Graduate School of Genome Science & Technology, University of Tennessee-Knoxville Knoxville (Knoxville, TN USA)

^d University of Sydney, School of Chemistry (Sydney, NSW, Australia)

^e Department of Biochemistry, Cellular, and Molecular Biology, University of Tennessee-Knoxville (Knoxville, TN USA)

^f Department of Electrical Engineering and Computer Science, University of Tennessee-Knoxville (Knoxville, TN USA)

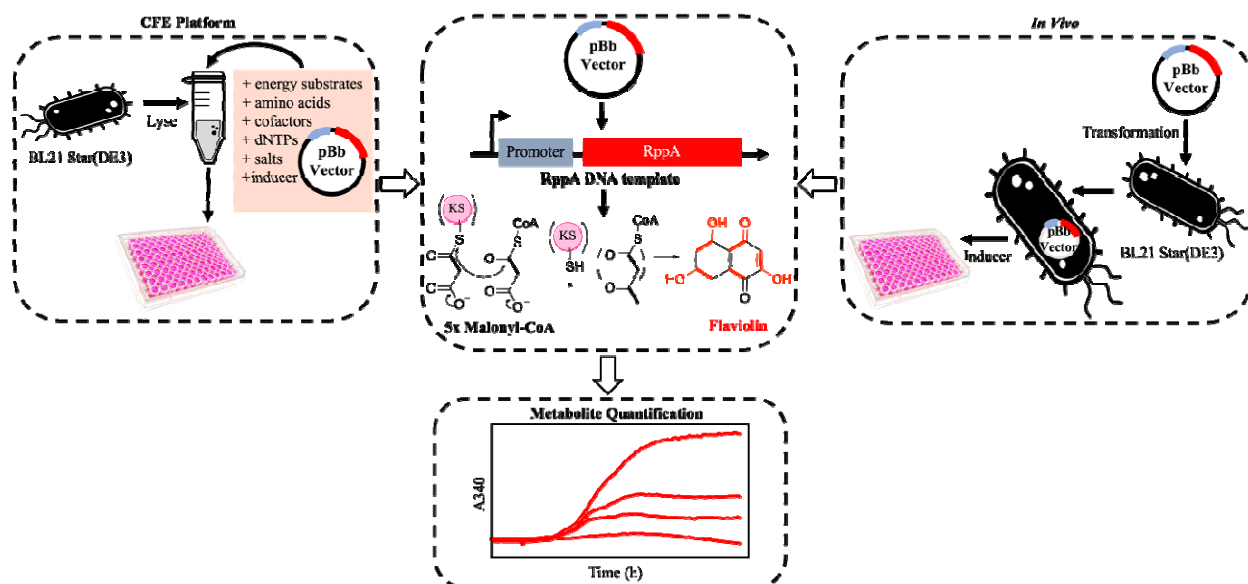
^g Department of Ecology & Evolutionary Biology, University of Tennessee-Knoxville (Knoxville, TN USA)

[‡] Authors Contributed Equally

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 1) a valuable expression platform for enzymes that are challenging to synthesize *in vivo*, and as 2) 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. We synergistically tune promoter and codon usage to improve flaviolin production from cell-free expressed RppA. We then 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. Refactoring promoters and/or coding sequences via CFE can be a valuable strategy to rapidly screen for catalytically functional production of enzymes from BCGs. By showing the coordinators between CFE versus *in vivo* expression, this work advances CFE as a tool for natural product research.

TOC



Introduction

Microbial secondary metabolism generates a vast number of complex secondary metabolites, or natural products, with varied chemistry.¹ 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.²⁻⁵ 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. 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.⁶ Where BGC expression in *E. coli* is possible, a myriad of parameters usually require optimization. Oftentimes, using the native coding sequence amplified from genome DNA with common expression systems (e.g., the pET expression system)^{7,8,9} results in poor expression or inclusion body formation.¹⁰ In other cases, 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 of 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* TX-TL PURE system and can bypass limitations in secondary metabolite production observed in *in vivo* expression systems.^{11,12} The TX-TL PURE system or PURExpress employ the minimal number of recombinant elements required for transcription and translation (TX-TL) 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.¹³⁻¹⁵ CFE systems are thus often used for synthesizing BGCs and their product metabolites, especially when cytotoxicity represents a limiting factor to heterologous *in vivo* production.¹⁶⁻¹⁸ 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 refactorizing strategies in a cell-free environment can benefit development of cell-free and cell-based BGC expression platforms.

While optimizations of protein expression via refactorizing have usually focused on robustly expressed reporters (e.g., sfGFP),¹⁹⁻²⁴ we sought to evaluate common refactorizing parameters using a reporter that was relevant to the enzymatic activity of genes involved in secondary metabolite formation.⁷ To further explore the ability to refactor for functional catalytic activity, we chose a model protein, RppA (40.1 kDa), which is a type III polyketide synthase from *Streptomyces griseus* that generates flaviolin, which is a red pigment that has limited catalytically functional expression in *E. coli* even with current optimization.²⁵⁻²⁷ 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,²⁸ and the pTet anhydrotetracycline promoter.²⁹ 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 refactorizing 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.

Results and discussion

Design of the vector system

To design a library of constructs to improve expression, we focused on two commonly varied refactorizing parameters: promoter choice and codon usage. We used four distinct coding sequences and four distinct promoters (**Figure 1A**). In terms of promoter choice, while the T7-

lac promoter³⁰ combination is the basis of the most commonly used expression system in *E. coli* (particularly the pET expression system)^{8,9} 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*.³⁰ Other commonly used vectors include the pTet promoter, which is anhydrotetracycline inducible,³⁰ and the pBAD promoter that is arabinose inducible²⁸ both of which are less leaky than the lac operon (**Figure 1B**). To obtain these constructs, we used a series of BioBrick vectors designed by Keasling and coworkers³⁰ 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.^{10,31–33} 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.^{34–40} While these genome and transcriptome optimized sequences can aid in successful expression of the heterologous product, improperly folded products can still result.^{41,42} Another approach is codon “harmonization,” which has been posited to improve expression via better co-translational folding.⁴³ Codon harmonization involves identifying and replicating patterns of codon usage in the donor organism with comparable patterns of codon usage in the heterologous host.^{44,45,44} 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 are more akin to how they originally evolved and therefore might result in proper protein folding.⁴⁵

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 (HC-*rppA*) and a new method based on ROC-SEMPPR (HR-*rppA*) (**Figure 1C**). Briefly, the CHARMING method applies a relative measure of codon usage called “% MinMax” (%MM).⁴² %MM values are computed as described by Chaney et al.^{33,46,47} 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/>).⁴⁸ 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 $|MM_{\text{original}} - MM_{\text{target}}|$ 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.⁴² 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.^{49,50} By simultaneously analyzing intragenic and intergenic patterns of synonymous codon usage within a genome, ROC-SEMPPR estimates differences in ribosome pausing times among synonymous codons translational efficiency and 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 genomes (in this case *E. coli*) 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 are described in **Figure 1**.

Figure 1.

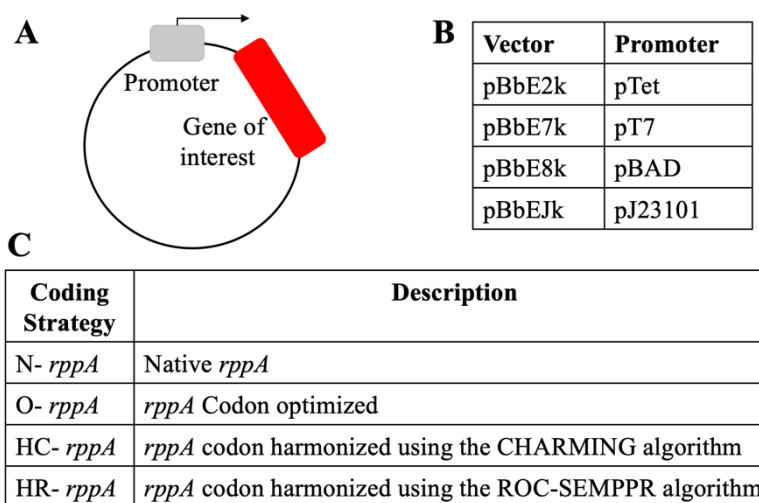


Figure 1. Design and nomenclature of plasmids in this study. (A) Each plasmid tested is composed of two modules: a backbone containing different promoters and the gene of interest (*rppA*) (B) Names of the BioBrick vectors (pBb) carrying different promoters.³⁰ (C) Names of the varied coding sequences.

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 pentaketide polyketide tetrahydroxynaphthalene (THN).^{25,26} Subsequently, THN undergoes a spontaneous oxidation reaction to convert THN to flaviolin (**Figure 2A**). The formation of the red-brown flaviolin pigment can thus be monitored as it readily absorbs light at 340 nm.²⁷ As metabolite production in an *E. coli* lysate-based cell-free system is correlated with the amount of protein that is expressed,⁵¹ 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 (**Figure S1**). 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 of a lacUV5 promoter.⁵² We have previously had success using this lysate for the production of pigments from *Streptomyces*.⁵¹ 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 sequences¹⁰ and 2) the T7 promoters (and specifically pET vectors) have been widely used in CFE.^{51,53–57} These initial experiments revealed superior pigment production at 30°C, so we proceeded with 30°C for all subsequent experiments (**Figure S2**). 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 2B**).

Figure 2.

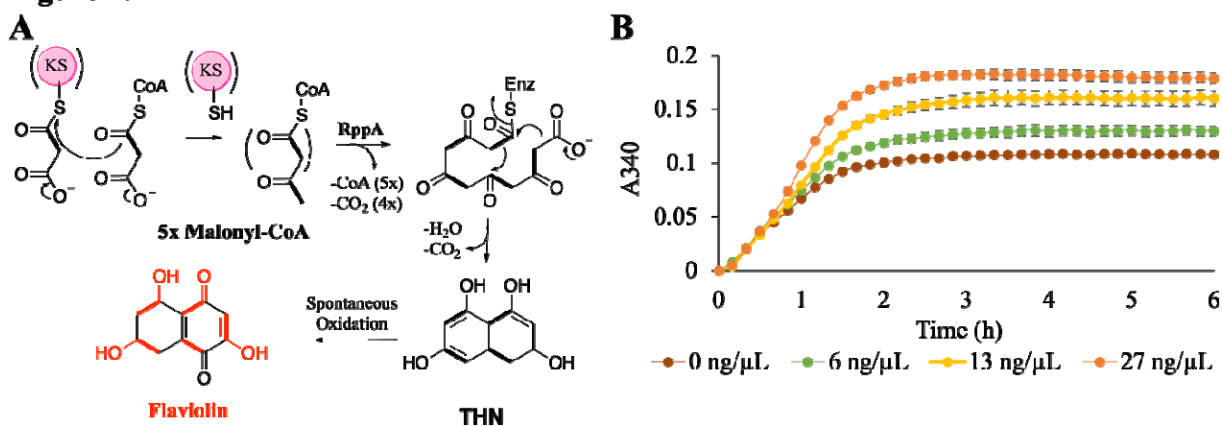


Figure 2. Flaviolin production via RppA in CFE reactions. (A) Schematic of flaviolin biosynthetic pathway by type III PKS, RppA. (B) CFE reactions were initiated with increasing concentrations of codon optimized *rppA* (*o-rppA*) driven by pT7 promoter in plasmid DNA, resulting in increasing flaviolin production (observed by increase of absorbance at 340 nm).

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

The IPTG 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.⁵⁸ However, its extreme promoter strength, combined with 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.^{56,59,60} Extensive efforts have been made to apply the T7 promoter series under the lac operator in CFE.^{53,61} However, other inducible systems remain extremely underexplored, with only a few reports of their usage in lysate based CFE systems.⁶²⁻⁶⁴

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/ μ L T7 RNA polymerase.⁶⁵ Maximal flaviolin production was observed with a concentration of 500 μ M IPTG (**Figure S3A, 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 μ M malonyl-CoA to the reactions enabled higher levels of flaviolin synthesis (**Figure S3B, C**). *E. coli* can possess low levels of malonyl-CoA endogenously,⁶⁶ 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 μ M) 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 (**Figures S4, S5**). For the pTet promoter, we observed the greatest expression level at 50 μ M anhydrotetracycline (**Figure S4A, B**), whereas 10 mM L-arabinose was greatest for the pBAD promoter (**Figure S5A, 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 (**Figure S6**). 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 (**Figure S7A-C**). We subsequently compared sfGFP synthesis from these inducible promoters to one another and to production from 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 (**Figure S7D**). While the constitutive promoter produced relatively low levels of sfGFP, it did result sfGFP expression over a faster timeframe compared to pTet and pBAD (**Figure S7D & E**). Thus, delayed CFE from pTet and pBAD is likely a result of their tighter regulation compared to pT7.

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 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 (**Figure S6**). 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);⁴² 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 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 3B**). Similarly, with the pBAD promoter, the native construct has low expression while HR-*rppA* has the best expression followed by HC-*rppA* (**Figure 3C**). Because these codon harmonization models do not account for inducible expression, we sought to

determine the effects of these re-coding strategies on flavin 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 (**Figure S7D**). Thus, these results also confirm that different proteins have varying optimal CFE expression conditions.⁵¹

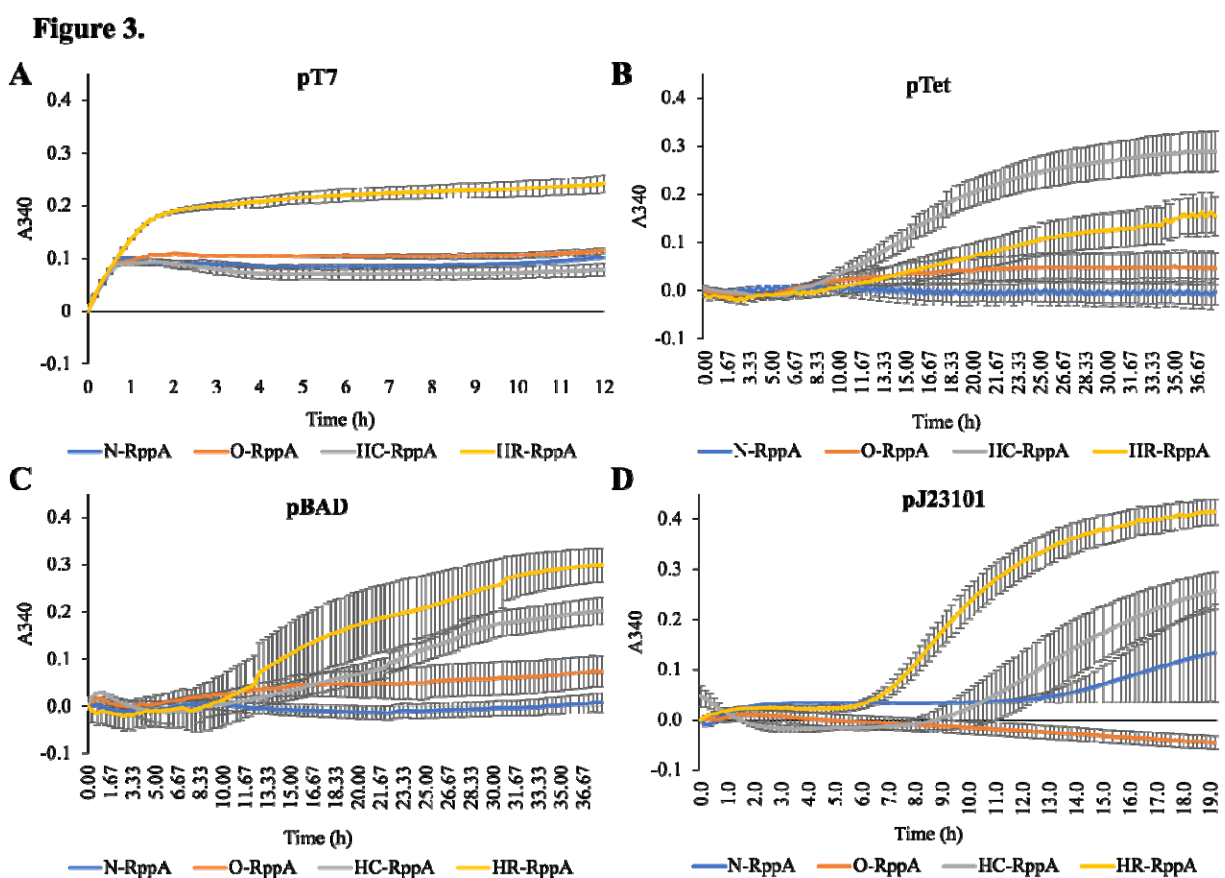


Figure 3. Flavin measurement in CFE reactions initiated with plasmids carrying different promoter and coding sequence combinations. Error bars represent the standard error of the mean ($n = 3$). (A) CFE reactions containing one of four pT7 promoter constructs, 50 ng/ μ L T7 RNA polymerase, and 500 μ M 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 μ M malonyl-CoA, and 50 μ M 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 μ M 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 μ M 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 ($n = 3$).

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*.^{54,67,68} 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, OD₆₀₀ of the codon-optimized construct of each promoter was measured to determine inducing time. Optimal OD₆₀₀ for each promoter/operator combination was based on literature precedent for standard ODs of induction for each promoter respectively (OD₆₀₀ = ~0.8 for T7-lac, OD₆₀₀ = ~0.2 for pBAD, OD₆₀₀ = ~0.6 for pTet).^{69–71} In a 96-well plate, from the initial culture with OD₆₀₀ = 0.05, pT7 constructs take 3.5 hr to reach OD₆₀₀ = 0.8, pTet constructs take 3 hours to reach OD₆₀₀ = 0.6, and pBAD constructs take 70 mins to reach OD₆₀₀ = 0.2. Next, we varied the inducer concentrations for each promoter - *rppA* DNA construct (**Figure S8**). Of the conditions we evaluated, we found that inducer concentrations are different with CFE reactions: IPTG concentration reaches greatest expression at 500 μ M, 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.³⁰

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 3A&D, 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 flavin synthesis from varying coding sequences under pBAD expression are indistinguishable *in vivo* (**Figure 4C**), and generally lower compared to production in the cell-free system (**Figure 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 flavin synthesis in pBAD-regulated CFE.²⁸ Notably, the current CFE system cannot be used to prototype promoter choice for the *in vivo* expression of RppA, given any coding sequence (**Figure 3 and Figure 4**). The pT7 expression measurements were collected when the inducible IPTG was added, which was after 3.5 hours of growth (**Figure 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 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* (**Figures S7D&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*.

Figure 4.

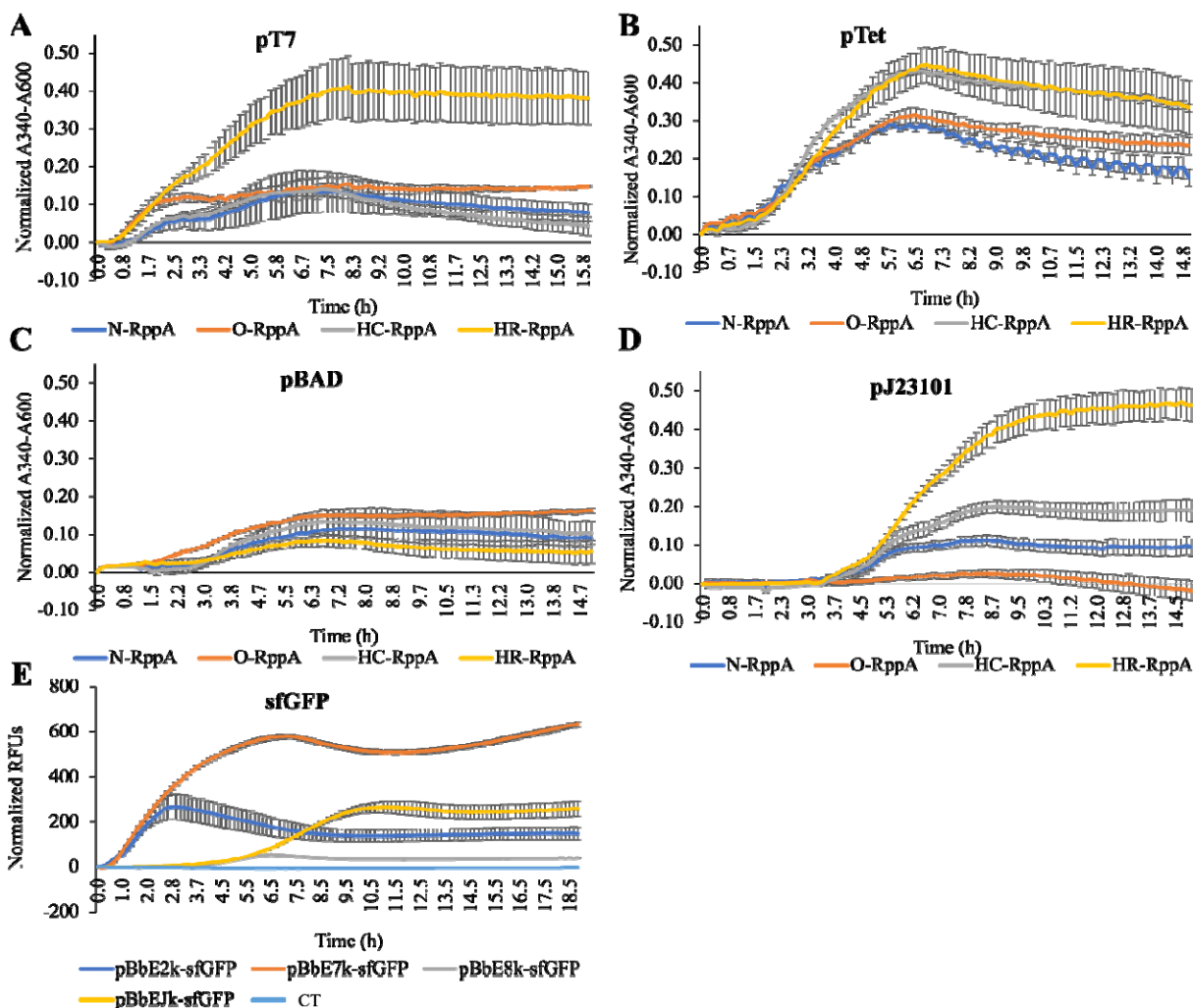


Figure 4. Flaviolin formation in *E. coli* BL21 (Star)DE3 cells carrying different promoter and coding sequence combinations. Flaviolin production in cells carrying (A) pT7 promoter constructs post-induction with 500 μ M IPTG (added after 3.5 hours of growth), (B) pTet promoter constructs post-induction with 100 nM anhydrotetracycline (added after 3 hours of growth), (C) pBAD constructs post-induction with 5000 μ M L-arabinose (added after 1 hour and 10 min), and (D) pJ23101 constructs. (E) Expression of BL21 Star(DE3) cells harboring sfGFP. Reactions were run in triplicate and read every 10 minutes for 20 hours. Error bars represent the standard error of the mean ($n = 3$).

Conclusion

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,⁷² 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.^{42,73} 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,⁷ 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,³³ 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.

Materials and methods

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 S1**). 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 S2**). 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 S3**.⁷⁴

***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 OD₆₀₀ = 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 A₃₄₀/A₆₀₀ were taken every 10 min for 20 hr.

***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 OD₆₀₀ 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 ¹H NMR: (400 MHz, acetone D₆) 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.^{75,76}

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 OD₆₀₀ ~ 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 μ L of 1 M DTT was added into each tube immediately after sonication. All samples were centrifuged at 12000 x g for 10 minutes at 4°C. The supernatant was collected without disturbing the pellet and centrifuged again to remove the remaining debris. The resulting supernatants were aliquoted into fresh centrifuge tubes, flash-frozen, and stored at -80°C.

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 μ g/mL folinic acid; 0.4 mM nicotinamide adenine dinucleotide (NAD), 0.27 mM coenzyme A (CoA), 4 mM oxalic acid, 1 mM, 1.5 mM spermidine, 57.33 mM HEPES buffer, 10 mM magnesium glutamate, 10 mM ammonium glutamate, 130 mM potassium glutamate, 2 mM each of the 20 amino acids, 33 mM phosphoenolpyruvate (PEP), 27 ng/ μ L DNA template and incubated at 30°C.⁶⁵ Reactions were set up in 10 μ L 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.

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 μ M 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. A₃₄₀ measurements were taken and normalized as described above.

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.

Supplementary data

Data availability

All the data described in this article are available within the article and in its online [supplementary files](#).

Funding

This work was supported by the University of Tennessee-Knoxville, the University of Tennessee-Oak Ridge Innovation Institute Science Alliance, the National Institutes of Health (R15GM145182), and the University of Sydney to C.B.B. C.B.B. is a member of the University of Sydney Drug Discovery Initiative, the University of Sydney Infectious Diseases Institute, and the University of Sydney Nanoscience Institute. This research was sponsored by the Genomic Science Program, US. Department of Energy, Office of Science, Biological, and Environmental Research as part of the Plant Microbe Interfaces Scientific Focus Area (<http://pmi.ornl.gov>). Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the Department of Energy under contract DE-AC05099OR2725. This manuscript has been authored by UT-Battelle, LLC under Contract DA-AC05-00OR2275 with the U.S. Department of Energy. The United States Government retains a nonexclusive, paid-up, irrevocable worldwide license to publish or reproduce the published form of this manuscript or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Access plan (<http://energy.gov/downloads/doe-public-access-plan>). J.L.N.D. and T.T.S. were supported by University of Tennessee-Oak Ridge Innovation Institute Science Alliance Graduate-Advancement, Training, and Education (GATE) fellowships. G.A., J.W.B., D.S.G., and E.G., were supported by the Advanced Undergraduate Research Activity (AURA) fellowships from the University of Tennessee-Knoxville Office of Undergraduate Research and Fellowships. M.S. was supported by a Summer Research Training (SmART) summer internship from the UT-Oak Ridge Innovation Institute (UT-ORII).

Conflict of interest statement.

All the authors declare that they have no competing interests.

References

- (1) Bérdy, J. (2005) Bioactive microbial metabolites. *J. Antibiot.* 58, 1–26.
- (2) Baltz, R. H. (2017) Gifted microbes for genome mining and natural product discovery. *J. Ind. Microbiol. Biotechnol.* 44, 573–588.

- (3) Palazzotto, E., Tong, Y., Lee, S. Y., and Weber, T. (2019) Synthetic biology and metabolic engineering of actinomycetes for natural product discovery. *Biotechnol. Adv.* 37, 107366.
- (4) Musiol-Kroll, E. M., Tocchetti, A., Sosio, M., and Stegmann, E. (2019) Challenges and advances in genetic manipulation of filamentous actinomycetes - the remarkable producers of specialized metabolites. *Nat. Prod. Rep.* 36, 1351–1369.
- (5) Drufva, E. E., Sword, T. T., and Bailey, C. B. (2022) Metabolic engineering of actinomycetes for natural product discovery, in *Natural Products from Actinomycetes: Diversity, Ecology and Drug Discovery* (Rai, R. V., and Bai, J. A., Eds.), pp 267–307. Springer Singapore, Singapore.
- (6) Stevens, D. C., Hari, T. P. A., and Boddy, C. N. (2013) The role of transcription in heterologous expression of polyketides in bacterial hosts. *Nat. Prod. Rep.* 30, 1391–1411.
- (7) Schmidt, M., Lee, N., Zhan, C., Roberts, J. B., Nava, A. A., Keiser, L. S., Vilchez, A. A., Chen, Y., Petzold, C. J., Haushalter, R. W., Blank, L. M., and Keasling, J. D. (2023) Maximizing heterologous expression of engineered type I polyketide synthases: investigating codon optimization strategies. *ACS Synth. Biol.*
- (8) Dubendorff, J. W., and Studier, F. W. (1991) Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* 219, 45–59.
- (9) William Studier, F., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) [6] Use of T7 RNA polymerase to direct expression of cloned genes, in *Gene Expression Technology*, pp 60–89. Elsevier.
- (10) 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 lavendulae* reveals insights on the effects of refactoring biosynthetic megasynthases for heterologous expression in *Escherichia coli*. *Protein Expr. Purif.* 210, 106317.
- (11) Wagner, L., Jules, M., and Borkowski, O. (2023) What remains from living cells in bacterial lysate-based cell-free systems. *Comput. Struct. Biotechnol. J.* 21, 3173–3182.
- (12) Tuckey, C., Asahara, H., Zhou, Y., and Chong, S. (2014) Protein synthesis using a reconstituted cell-free system. *Curr. Protoc. Mol. Biol.* 108, 16.31.1-16.31.22.
- (13) Dinglasan, J. L. N., and Doktycz, M. J. (2023) Rewiring cell-free metabolic flux in *E. coli* Lysates using a Block-Push-Pull Approach. *Synth. Biol.*
- (14) Garcia, D. C., Dinglasan, J. L. N., Shrestha, H., Abraham, P. E., Hettich, R. L., and Doktycz, M. J. (2021) A lysate proteome engineering strategy for enhancing cell-free metabolite production. *Metab. Eng. Commun.* 12, e00162.

- (15) Dinglasan, J. L. N., Reeves, D. T., Hettich, R. L., and Doktycz, M. J. (2021) Liquid Chromatography coupled to refractive index or mass spectrometric detection for metabolite profiling in lysate-based cell-free systems. *J. Vis. Exp.*
- (16) Mouncey, N. J., Otani, H., Udvary, D., and Yoshikuni, Y. (2019) New voyages to explore the natural product galaxy. *J. Ind. Microbiol. Biotechnol.* 46, 273–279.
- (17) Bogart, J. W., Cabezas, M. D., Vögeli, B., Wong, D. A., Karim, A. S., and Jewett, M. C. (2021) Cell-free exploration of the natural product chemical space. *Chembiochem* 22, 84–91.
- (18) Ji, X., Liu, W.-Q., and Li, J. (2022) Recent advances in applying cell-free systems for high-value and complex natural product biosynthesis. *Curr. Opin. Microbiol.* 67, 102142.
- (19) Garenne, D., Haines, M. C., Romantseva, E. F., Freemont, P., Strychalski, E. A., and Noireaux, V. (2021) Cell-free gene expression. *Nat. Rev. Methods Primers* 1, 49.
- (20) Moore, S. J., MacDonald, J. T., Wienecke, S., Ishwarbhai, A., Tsipa, A., Aw, R., Kylilis, N., Bell, D. J., McClymont, D. W., Jensen, K., Polizzi, K. M., Biedendieck, R., and Freemont, P. S. (2018) Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria. *Proc Natl Acad Sci USA* 115, E4340–E4349.
- (21) Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* 24, 79–88.
- (22) Lentini, R., Forlin, M., Martini, L., Del Bianco, C., Spencer, A. C., Torino, D., and Mansy, S. S. (2013) Fluorescent proteins and in vitro genetic organization for cell-free synthetic biology. *ACS Synth. Biol.* 2, 482–489.
- (23) Jew, K., Smith, P. E. J., So, B., Kasman, J., Oza, J. P., and Black, M. W. (2022) Characterizing and Improving pET Vectors for Cell-free Expression. *Front. Bioeng. Biotechnol.* 10, 895069.
- (24) Burrington, L. R., Watts, K. R., and Oza, J. P. (2021) Characterizing and Improving Reaction times for E. coli-based cell-free protein synthesis. *ACS Synth. Biol.* 10, 1821–1829.
- (25) Funai, N., Ohnishi, Y., Fujii, I., Shibuya, M., Ebizuka, Y., and Horinouchi, S. (1999) A new pathway for polyketide synthesis in microorganisms. *Nature* 400, 897–899.
- (26) Funai, N., Ohnishi, Y., Ebizuka, Y., and Horinouchi, S. (2002) Properties and substrate specificity of RppA, a chalcone synthase-related polyketide synthase in *Streptomyces griseus*. *J. Biol. Chem.* 277, 4628–4635.
- (27) Yang, D., Kim, W. J., Yoo, S. M., Choi, J. H., Ha, S. H., Lee, M. H., and Lee, S. Y. (2018) Repurposing type III polyketide synthase as a malonyl-CoA biosensor for metabolic engineering in bacteria. *Proc Natl Acad Sci USA* 115, 9835–9844.

- (28) Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* *177*, 4121–4130.
- (29) Lee, S. K., Newman, J. D., and Keasling, J. D. (2005) Catabolite repression of the propionate catabolic genes in *Escherichia coli* and *Salmonella enterica*: evidence for involvement of the cyclic AMP receptor protein. *J. Bacteriol.* *187*, 2793–2800.
- (30) Lee, T. S., Krupa, R. A., Zhang, F., Hajimorad, M., Holtz, W. J., Prasad, N., Lee, S. K., and Keasling, J. D. (2011) BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J. Biol. Eng.* *5*, 12.
- (31) Krefft, D., Papkov, A., Zylicz-Stachula, A., and Skowron, P. M. (2017) 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* *12*, e0186633.
- (32) Walsh, I. M., Bowman, M. A., Soto Santarriaga, I. F., Rodriguez, A., and Clark, P. L. (2020) Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. *Proc Natl Acad Sci USA* *117*, 3528–3534.
- (33) Chaney, J. L., Steele, A., Carmichael, R., Rodriguez, A., Specht, A. T., Ngo, K., Li, J., Emrich, S., and Clark, P. L. (2017) Widespread position-specific conservation of synonymous rare codons within coding sequences. *PLoS Comput. Biol.* *13*, e1005531.
- (34) Welch, M., Govindarajan, S., Ness, J. E., Villalobos, A., Gurney, A., Minshull, J., and Gustafsson, C. (2009) Design parameters to control synthetic gene expression in *Escherichia coli*. *PLoS ONE* *4*, e7002.
- (35) Mellitzer, A., Weis, R., Glieder, A., and Flicker, K. (2012) Expression of lignocellulolytic enzymes in *Pichia pastoris*. *Microb. Cell Fact.* *11*, 61.
- (36) Kodumal, S. J., Patel, K. G., Reid, R., Menzella, H. G., Welch, M., and Santi, D. V. (2004) Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc Natl Acad Sci USA* *101*, 15573–15578.
- (37) Feng, Z., Zhang, L., Han, X., and Zhang, Y. (2010) Codon optimization of the calf prochymosin gene and its expression in *Kluyveromyces lactis*. *World J. Microbiol. Biotechnol.* *26*, 895–901.
- (38) Marlatt, N. M., Spratt, D. E., and Shaw, G. S. (2010) Codon optimization for enhanced *Escherichia coli* expression of human S100A11 and S100A1 proteins. *Protein Expr. Purif.* *73*, 58–64.

- (39) Villalobos, A., Ness, J. E., Gustafsson, C., Minshull, J., and Govindarajan, S. (2006) Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* 7, 285.
- (40) Richardson, S. M., Wheelan, S. J., Yarrington, R. M., and Boeke, J. D. (2006) GeneDesign: rapid, automated design of multikilobase synthetic genes. *Genome Res.* 16, 550–556.
- (41) Mignon, C., Mariano, N., Stadthagen, G., Lugari, A., Lagoutte, P., Donnat, S., Chenavas, S., Perot, C., Sodoyer, R., and Werle, B. (2018) Codon harmonization - going beyond the speed limit for protein expression. *FEBS Lett.* 592, 1554–1564.
- (42) Wright, G., Rodriguez, A., Li, J., Milenkovic, T., Emrich, S. J., and Clark, P. L. (2022) CHARMING: Harmonizing synonymous codon usage to replicate a desired codon usage pattern. *Protein Sci.* 31, 221–231.
- (43) Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J. M., Sachs, M. S., and Liu, Y. (2013) Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* 495, 111–115.
- (44) Willems, T., Hectors, W., Rombaut, J., De Rop, A.-S., Goegebeur, S., Delmulle, T., De Mol, M. L., De Maeseneire, S. L., and Soetaert, W. K. (2023) An exploratory in silico comparison of open-source codon harmonization tools. *Microb. Cell Fact.* 22, 227.
- (45) Shabalina, S. A., Spiridonov, N. A., and Kashina, A. (2013) Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. *Nucleic Acids Res.* 41, 2073–2094.
- (46) Clarke, T. F., and Clark, P. L. (2008) Rare codons cluster. *PLoS ONE* 3, e3412.
- (47) Rodriguez, A., Wright, G., Emrich, S., and Clark, P. L. (2018) %MinMax: A versatile tool for calculating and comparing synonymous codon usage and its impact on protein folding. *Protein Sci.* 27, 356–362.
- (48) Nakamura, Y., Gojobori, T., and Ikemura, T. (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* 28, 292.
- (49) Gilchrist, M. A., Chen, W.-C., Shah, P., Landerer, C. L., and Zaretzki, R. (2015) Estimating Gene Expression and Codon-Specific Translational Efficiencies, Mutation Biases, and Selection Coefficients from Genomic Data Alone. *Genome Biol. Evol.* 7, 1559–1579.
- (50) Cope, A. L., and Gilchrist, M. A. (2022) Quantifying shifts in natural selection on codon usage between protein regions: a population genetics approach. *BMC Genomics* 23, 408.
- (51) 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.

- (52) McKevitt, M., Patel, K., Smajs, D., Marsh, M., McLoughlin, M., Norris, S. J., Weinstock, G. M., and Palzkill, T. (2003) Systematic cloning of *Treponema pallidum* open reading frames for protein expression and antigen discovery. *Genome Res.* *13*, 1665–1674.
- (53) Senda, N., Enomoto, T., Kihara, K., Yamashiro, N., Takagi, N., Kiga, D., and Nishida, H. (2022) Development of an expression-tunable multiple protein synthesis system in cell-free reactions using T7-promoter-variant series. *Synth Biol (Oxf)* *7*, ysac029.
- (54) Karim, A. S., Liew, F. E., Garg, S., Vögeli, B., Rasor, B. J., Gonnot, A., Pavan, M., Juminaga, A., Simpson, S. D., Köpke, M., and Jewett, M. C. (2020) Modular cell-free expression plasmids to accelerate biological design in cells. *Synth Biol (Oxf)* *5*, ysaa019.
- (55) Swartz, J. R., Jewett, M. C., and Woodrow, K. A. (2004) Cell-free protein synthesis with prokaryotic combined transcription-translation. *Methods Mol. Biol.* *267*, 169–182.
- (56) Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., and Noireaux, V. (2013) Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp.* e50762.
- (57) Garenne, D., Thompson, S., Brisson, A., Khakimzhan, A., and Noireaux, V. (2021) The all-E. coliTXTL toolbox 3.0: new capabilities of a cell-free synthetic biology platform. *Synth Biol (Oxf)* *6*, ysab017.
- (58) Tokmakov, A. A., and Fukami, Y. (2010) Activation of T7 RNA polymerase in *Xenopus* oocytes and cell-free extracts. *Genes Cells* *15*, 1136–1144.
- (59) Hagen, A., Poust, S., de Rond, T., Yuzawa, S., Katz, L., Adams, P. D., Petzold, C. J., and Keasling, J. D. (2014) In vitro analysis of carboxyacyl substrate tolerance in the loading and first extension modules of borrelidin polyketide synthase. *Biochemistry* *53*, 5975–5977.
- (60) Hagen, A., Poust, S., Rond, T. de, Fortman, J. L., Katz, L., Petzold, C. J., and Keasling, J. D. (2016) Engineering a polyketide synthase for in vitro production of adipic acid. *ACS Synth. Biol.* *5*, 21–27.
- (61) Karig, D. K., Iyer, S., Simpson, M. L., and Doktycz, M. J. (2012) Expression optimization and synthetic gene networks in cell-free systems. *Nucleic Acids Res.* *40*, 3763–3774.
- (62) Borkowski, O., Bricio, C., Murgiano, M., Rothschild-Mancinelli, B., Stan, G.-B., and Ellis, T. (2018) Cell-free prediction of protein expression costs for growing cells. *Nat. Commun.* *9*, 1457.
- (63) Brooks, R., Morici, L., and Sandoval, N. (2023) Cell Free Bacteriophage Synthesis from Engineered Strains Improves Yield. *ACS Synth. Biol.* *12*, 2418–2431.
- (64) Guo, S., and Murray, R. M. (2019) Construction of Incoherent Feedforward Loop Circuits in a Cell-Free System and in Cells. *ACS Synth. Biol.* *8*, 606–610.

- (65) Levine, M. Z., Gregorio, N. E., Jewett, M. C., Watts, K. R., and Oza, J. P. (2019) Escherichia coli-Based Cell-Free Protein Synthesis: Protocols for a robust, flexible, and accessible platform technology. *J. Vis. Exp.*
- (66) Takamura, Y., and Nomura, G. (1988) 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.* 134, 2249–2253.
- (67) Karim, A. S., Dudley, Q. M., Juminaga, A., Yuan, Y., Crowe, S. A., Heggestad, J. T., Garg, S., Abdalla, T., Grubbe, W. S., Rasor, B. J., Coar, D. N., Torculas, M., Krein, M., Liew, F. E., Quattlebaum, A., Jensen, R. O., Stuart, J. A., Simpson, S. D., Köpke, M., and Jewett, M. C. (2020) In vitro prototyping and rapid optimization of biosynthetic enzymes for cell design. *Nat. Chem. Biol.* 16, 912–919.
- (68) Vögeli, B., Schulz, L., Garg, S., Tarasava, K., Clomburg, J. M., Lee, S. H., Gonnot, A., Mouly, E. H., Kimmel, B. R., Tran, L., Zeleznik, H., Brown, S. D., Simpson, S. D., Mrksich, M., Karim, A. S., Gonzalez, R., Köpke, M., and Jewett, M. C. (2022) Cell-free prototyping enables implementation of optimized reverse β -oxidation pathways in heterotrophic and autotrophic bacteria. *Nat. Commun.* 13, 3058.
- (69) Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J., and Li, Q. (2009) Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Sci.* 18, 936–948.
- (70) Geurink, P. P., van der Heden van Noort, G. J., Mulder, M. P. C., Knaap, R. C. M., Kikkert, M., and Ovaa, H. (2019) Profiling DUBs and Ubl-specific proteases with activity-based probes. *Meth. Enzymol.* 618, 357–387.
- (71) Khlebnikov, A., Risa, O., Skaug, T., Carrier, T. A., and Keasling, J. D. (2000) Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J. Bacteriol.* 182, 7029–7034.
- (72) Chappell, J., Jensen, K., and Freemont, P. S. (2013) Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. *Nucleic Acids Res.* 41, 3471–3481.
- (73) Wright, G., Rodriguez, A., Clark, P. L., and Emrich, S. (2019) A new look at codon usage and protein expression. *Epic Ser. Comput.* 60, 104–112.
- (74) Hillson, N. J., Rosengarten, R. D., and Keasling, J. D. (2012) j5 DNA assembly design automation software. *ACS Synth. Biol.* 1, 14–21.
- (75) Medina, R., Lucentini, C. G., Franco, M. E. E., Petroselli, G., Rosso, J. A., Erra-Balsells, R., Balatti, P. A., and Saparrat, M. C. N. (2018) Identification of an intermediate for 1,8-dihydroxynaphthalene-melanin synthesis in a race-2 isolate of *Fulvia fulva* (syn. *Cladosporium fulvum*). *Heliyon* 4, e01036.

- (76) Becker, K., Lambert, C., Wieschhaus, J., and Stadler, M. (2020) Phylogenetic Assignment of the Fungicolous *Hypoxylon invadens* (Ascomycota, Xylariales) and Investigation of its Secondary Metabolites. *Microorganisms* 8.