scientific reports

OPEN

Profiling expression strategies for a type III polyketide synthase in a lysate-based, cell-free system

Tien T. Sword[1,8], Jaime Lorenzo N. Dinglasan[2,3,8], Ghaeath S. K. Abbas[1,4], J. William Barker[1], Madeline E. Spradley[5], Elijah R. Greene[1], Damian S. Gooden[1], Scott J. Emrich[3,6,7], Michael A. Gilchrist[3,7], Mitchel J. Doktycz[2,3✉] & Constance B. Bailey[1,3,4✉]

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* often results in a myriad of unpredictable issues with regard to 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 can be a lengthy trial-and-error process. Cell-free expression (CFE) has emerged as a valuable expression platform as a testbed for rapid prototyping expression parameters. 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 applied a library of constructs with different combinations of promoters and *rppA* coding sequences to investigate the synergies 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. The promoter strategy when applied to RppA was not completely correlated with that observed with GFP, indicating that different promoter strategies should be applied for different proteins. In vivo experiments suggest that there is correlation, but not complete alignment between expressing in cell free and in vivo. 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.

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 often spend substantial time optimizing soluble, appropriately folded functional expression, presenting a bottleneck to enzymatic 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

[1]Department of Chemistry, University of Tennessee-Knoxville, Knoxville, TN, USA. [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA. [3]Graduate School of Genome Science and Technology, University of Tennessee-Knoxville, Knoxville, TN, USA. [4]School of Chemistry, University of Sydney, Sydney, NSW, Australia. [5]Department of Biochemistry, Cellular, and Molecular Biology, University of Tennessee-Knoxville, Knoxville, TN, USA. [6]Department of Electrical Engineering and Computer Science, University of Tennessee-Knoxville, Knoxville, TN, USA. [7]Department of Ecology and Evolutionary Biology, University of Tennessee-Knoxville, Knoxville, TN, USA. [8]These authors contributed equally: Tien T. Sword and Jaime Lorenzo N. Dinglasan. ✉email: doktyczmj@ornl.gov; constance.bailey@sydney.edu.au

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. 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[7,8]. 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[9–11]. CFE systems thus present an emerging alternative approach for synthesizing BGCs and their product metabolites, especially when cytotoxicity represents a limiting factor to heterologous in vivo production[12–14]. 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)[15–20], we sought to evaluate common refactoring parameters using a reporter that was relevant to the enzymatic activity of genes involved in secondary metabolite formation[21]. 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[22–24]. RppA is a type III polyketide synthase from *Streptomyces griseus*. As a type III PKS (as opposed to the type I and type II PKSs), it condenses free malonyl CoA directly as opposed to covalently transferring it to a carrier protein, thus no phosphopantetheinyl transferase is required[25] for activation to a holo protein. Applying flaviolin production as a reporter for catalytically functional RppA expression, we varied parameters that are relevant to improving catalytically functional expression. This included the use of three different, commonly used inducible promoters: the workhorse T7/*lac* system, the pBAD arabinose promoter[26], and the pTet anhydrotetracycline promoter[27]. 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. 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 (Fig. 1).
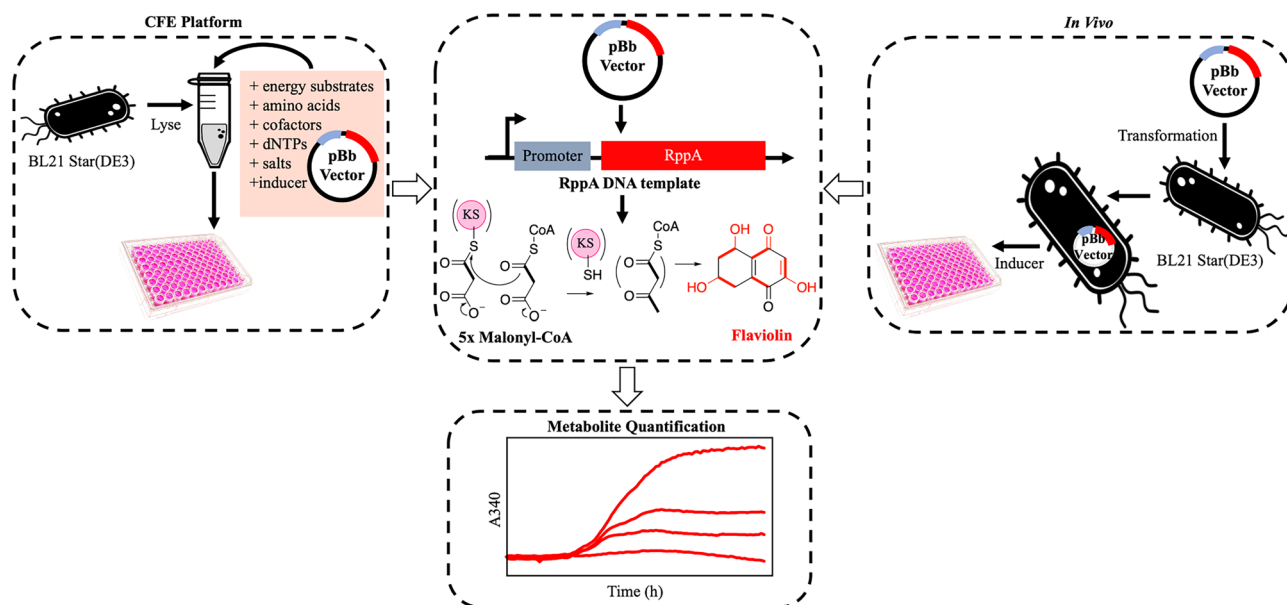

Figure 1. Overall schematic for monitoring and prototyping RppA activity.

Results and discussion
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. The choice of these two key parameters were rationalized because these are two of the most common alterations when trying to troubleshoot soluble expression of high GC in *E. coli*[21,28]. We used four distinct coding sequences and four distinct promoters (Fig. 2A). In terms of promoter choice, while the T7-*lac* promoter[28] combination is the basis of the most commonly used expression system in *E. coli* (particularly the pET expression system)[29,30] 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[28]. Other commonly used vectors include the pTet promoter, which is anhydrotetracycline inducible[28], and the pBAD promoter that is arabinose inducible[26] both of which are less leaky than the *lac* operon (Fig. 2B). To obtain these constructs, we used a series of BioBrick vectors designed by Keasling and coworkers[28] 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[31–34]. 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[35–41]. While these genome and transcriptome-optimized sequences can aid in the successful expression of the heterologous product, improperly folded products can still result[42,43]. Another approach is codon "harmonization," which has been posited to improve expression via better co-translational folding[44]. Codon harmonization involves identifying and replicating patterns of codon usage in the donor organism with comparable patterns of codon usage in the heterologous host[45,46]. 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[47].

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*)[43] and a new method based on ribosome overhead costs, Stochastic Evolutionary Model of Protein Production Rate (ROC-SEMPPR)[48] (HR-*rppA*) (Fig. 2C). Briefly, the CHARMING method applies a relative measure of codon usage called "% Min-Max" (%MM)[43]. %MM values are computed as described by Chaney et al.[34,49,50] 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/)[51]. 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_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[43]. 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,
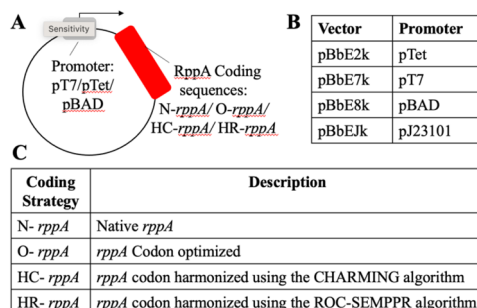

A		**B**	

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 2. 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[28]. (**C**) Names of the varied coding sequences.

population genetics-based model of sequence evolution, which includes the contributions of selection, mutation bias, and genetic drift, to an organism's coding sequences[48,52]. 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 Fig. 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 malonyl-CoA, resulting the pentaketide tetrahydroxynapthalene (THN)[22,23]. Subsequently, THN undergoes a spontaneous oxidation reaction to convert THN to flaviolin (Fig. 3A). The formation of the red-brown flaviolin pigment can thus be monitored as it readily absorbs light at 340 nm above cellular background[24,53]. As metabolite production in an *E. coli* lysate-based cell-free system is correlated with the amount of protein that is expressed[54], 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 (Fig. 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) strain that has the DE3 lysogen under control of a *lac*-UV5 promoter[55]. We have previously had success using this lysate for the production of pigments from *Streptomyces*[54]. 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[33] and (2) the T7 promoters (and specifically pET vectors) have been widely used in CFE[54,56–60]. These initial experiments revealed superior pigment production at 30 °C, so we proceeded with 30 °C for all subsequent experiments (Fig. 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 (Fig. 3B). We found that flaviolin signal reached its maximal value after approximately 2–3 h despite the usage of a modified PANOx-SP system[61] which typically prolongs protein synthesis up to ~10 h. We are unsure of the major limiting factor resulting in flaviolin production which could related to enzymatic activity or substrate concentration. This also illustrates the complexity of monitoring an enzymatic reporter (as opposed to a reporter for protein synthesis such as sfGFP) as reaction kinetics, substrate turnover, and degradation can all come into play when monitoring output.
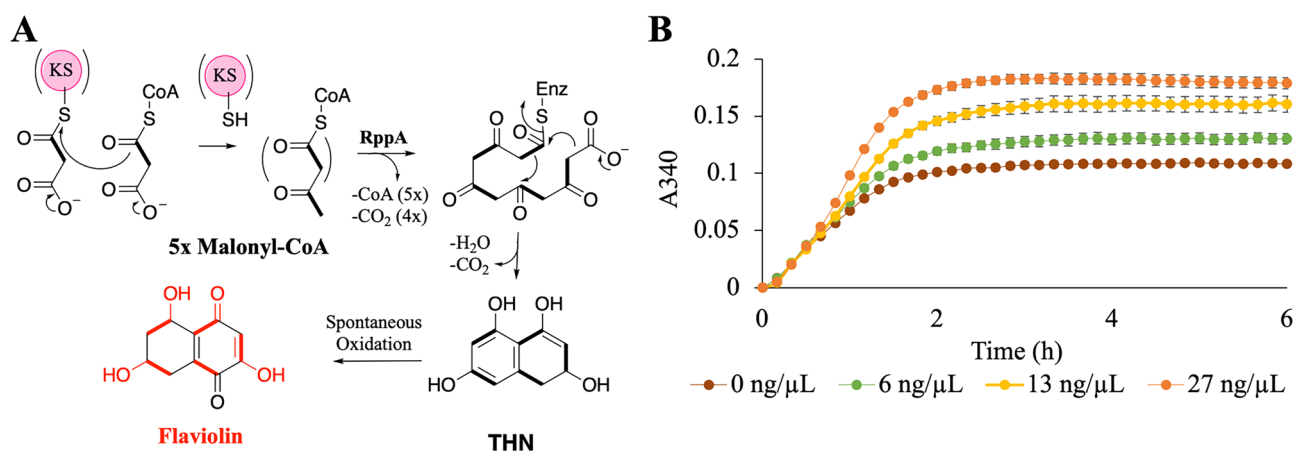

Figure 3. Flaviolin production via RppA in CFE reactions. (**A**) Schematic of flaviolin biosynthetic pathway by type III PKS, RppA. (**B**) CFE reactions were initiated with increasing concentrations of codon optimized *rppA* (O-*rppA*) driven by pT7 promoter in plasmid DNA, resulting in increasing flaviolin production (observed by increase of absorbance at 340 nm).

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

The 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*[61]. 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 the expression of proteins from *Streptomyces* that don't express under standard expression conditions (e.g. commercial pET vectors with a T7-lac system). For example, 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[59,62,63]. Extensive efforts have been made to apply the T7 promoter series under the *lac* operator in CFE[56,64]. However, other inducible systems remain extremely underexplored, with only a few reports of their usage in lysate-based CFE systems[65–67].

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[68]. Maximal flaviolin production was observed with a concentration of 500 μM IPTG (Fig. 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 ʟ-arabinose, respectively. Initial efforts did not result in detectable flaviolin production. We hypothesized that flaviolin signals are not detectable under these conditions due to lower soluble protein expression levels (further supported by Western blot, Figs. S9, S11), and, consequently, less flux being driven from endogenous malonyl-CoA precursor pools to flaviolin formation.

Thus, we hypothesized that detectability would be improved if we increased malonyl-CoA substrate availability. To test this, malonyl-CoA was added to reactions containing pBbE7k-O-*rppA* plasmid DNA at concentrations between 0 and 5000 μM (Fig. S3B). Indeed, adding up to 500 μM malonyl-CoA resulted in higher levels of flaviolin synthesis (Fig. S3B, D) and consequentially afforded successful detection of flaviolin from constructs driven by pBAD and pTet. A wide range of anhydrotetracycline and ʟ-arabinose concentrations were tested using pBbE2k-O-*rppA* (O-*rppA* driven by pTet) and pBbE8k-O-*rppA* (O-*rppA* driven by pBAD), respectively (Figs. S4, S5). For the pTet promoter, we observed the greatest expression level at 50 μM anhydrotetracycline (Fig. S4A,B), whereas 10 mM ʟ-arabinose was greatest for the pBAD promoter (Fig. 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 (Fig. 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 (Fig. S7A–C). We subsequently compared sfGFP synthesis from these inducible promoters and a constitutive promoter, pJ23101, in an analogous biobrick vector (pBbEJk). Evidently, there is a delay in the expression of sfGFP under the control of either pTet, pBAD, and pJ23101 promoters compared to pT7 (Fig. S7D). 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 (Fig. 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 (Fig. 4A–C). Whether or not ROC-SEMPPR will perform similarly well in other situations remains to be determined. It does, however, suggest that simply 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 which was the original rationale for developing the ROC-SEMPRR model. In addition, we observed that expression with the pT7 promoter is drastically increased when using the lysate that contained endogenous IPTG compared to lysate with exogenous IPTG (Fig. 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)[43]; however, because few rare
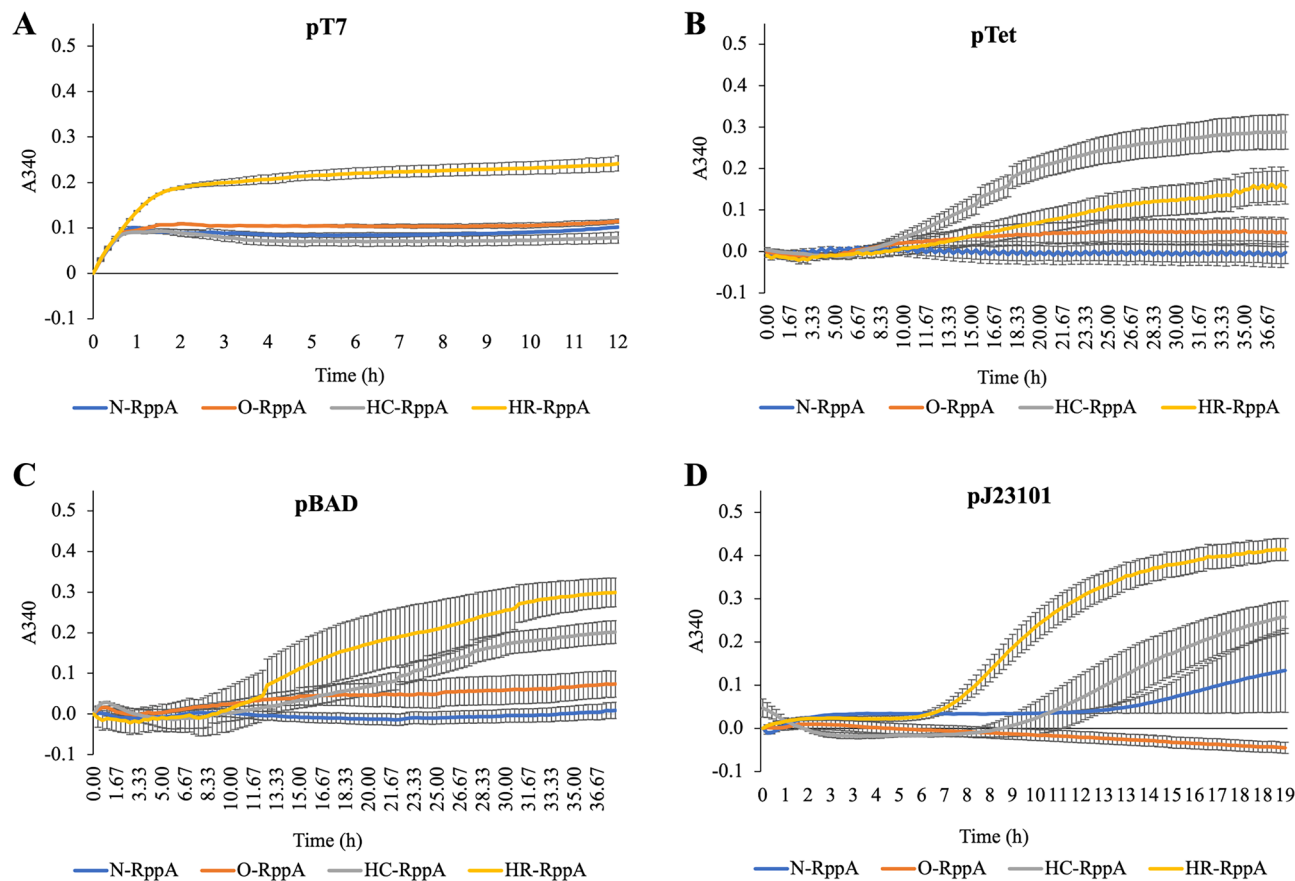
Figure 4. Flaviolin measurement in CFE reactions initiated with plasmids carrying different promoter and coding sequence combinations. Error bars represent the standard error of the mean ($n = 3$). (**A**) CFE reactions containing one of four pT7 promoter constructs, 50 ng/µL T7 RNA polymerase, and 500 µM IPTG. Reactions were run in triplicate and read every 10 min for 12 h. (**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 min for 38 h. (**C**) CFE reaction containing four pBAD promoter constructs plasmid DNA, 500 µM malonyl-CoA, and 10 mM ʟ-arabinose. Reactions were run in triplicate and read every 20 min for 38 h. (**D**) CFE reaction containing four pJ23101 promoter constructs plasmid DNA and 500 µM malonyl-CoA. Reactions were run in triplicate and read every 10 min for 20 h. All reactions were prepared with *E. coli* BL21 Star(DE3) lysates. Error bars represent the standard error of the mean ($n = 3$).

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 (Fig. 4A). 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 (Fig. 4B). Similarly, with the pBAD promoter, the native construct has low expression while HR-*rppA* has the best expression followed by HC-*rppA* (Fig. 4C). Due to the delay in expression observed while optimizing the induction of pTet and pBAD (Figs. S4, and S5), we monitored RppA production for a longer time period compared to the T7 constructs. Because these codon harmonization models do not account for inducible expression, 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 (Fig. 4D). 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. The level of protein expressed is correlated with the amount of flaviolin detected and can be visualized by western blot analysis of pooled lysate reactions (Figs. S9, S11). Additionally, unlike RppA, sfGFP expressed best under pT7 control while the constitutive promoter does not express well in these CFE conditions (Fig. S7). Thus, these results also confirm that different proteins have varying optimal CFE expression conditions[54].

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[57,69,70]. 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)[71–73]. In a 96-well plate, from the initial culture with OD_{600} = 0.05, pT7 constructs take 3.5 h to reach OD_{600} = 0.8, pTet constructs take 3 h to reach OD_{600} = 0.6, and pBAD constructs take 70 min to reach OD_{600} = 0.2. Next, we varied the inducer concentrations for each promoter (Fig. S8). 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 ʟ-arabinose at 5 mM. These conditions correlate with previous reported for RFP, thus, they were used for sfGFP expression[28].

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 (Fig. 5A,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 vivo (Fig. 5C), and generally lower compared to production in the cell-free system (Fig. 4C). Efficient catabolism of ʟ-arabinose by *E. coli* cells is a drawback of
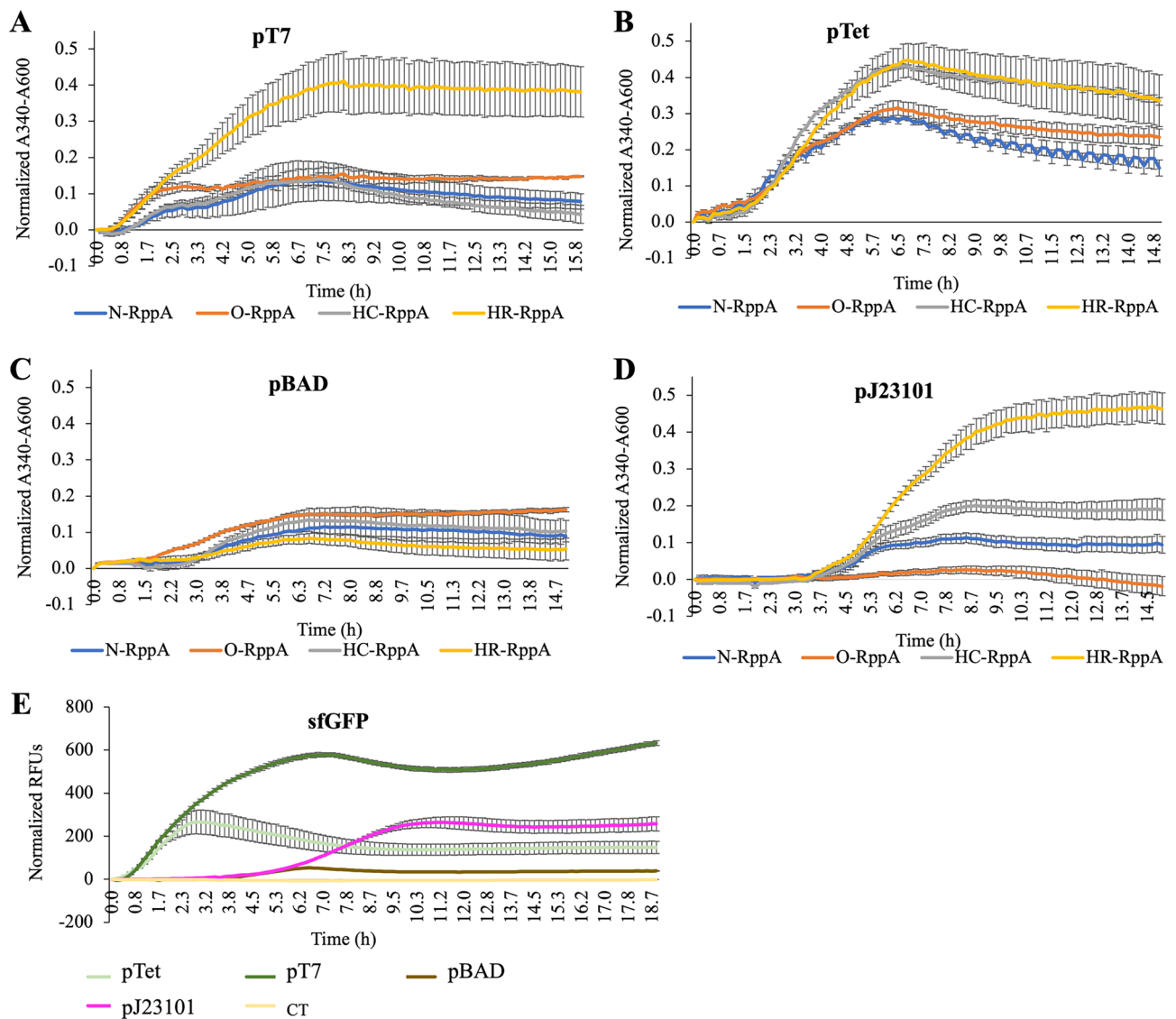

Figure 5. Flaviolin formation in *E. coli* BL21 (Star)DE3 cells carrying different promoter and coding sequence combinations. Flaviolin production in cells carrying (**A**) pT7 promoter constructs post-induction with 500 µM IPTG (added after 3.5 h of growth), (**B**) pTet promoter constructs post-induction with100 nM anhydrotetracycline (added after 3 h of growth), (**C**) pBAD constructs post-induction with 5000 µM ʟ-arabinose (added after 1 h 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 min for 20 h. Error bars represent the standard error of the mean (n = 3).

arabinose-inducible promoters, which may be a potential cause for better flaviolin synthesis in pBAD-regulated CFE[26]. Notably, the current CFE system cannot be used to prototype promoter choice for the in vivo expression of RppA, given any coding sequence (Figs. 4, 5). The pT7 expression measurements were collected when the inducible IPTG was added, which was after 3.5 h of growth (Fig. 5A). While the measurement of the pJ23101 constructs were collected right after inoculation into the 96-well plate. It takes about 3–4 h for the cells to grow before they start to produce pigment. This explains the delay occurring in the pJ23101 promoter (Fig. 5D). The level of protein expressed correlates with the amount of flaviolin detected and can be visualized by western blot analysis of pooled lysate reactions (Figs. S10, S11). 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 (Figs. 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.

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[74], 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[43,75]. 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 an actinomycete[21], 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[34], 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

Escherichia 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-terminal strep tag, TGGAGCCATCCGCAGTTCGAA AAA, 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[76].

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 h.

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 h, 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 h. The culture was then centrifuged at $5000 \times g$ for 30 min. The pink supernatant was adjusted to pH 2 with 3 M HCl and incubated at 4 °C overnight to precipitate flaviolin. Pigments were recovered by centrifugation at $5000 \times g$ for 30 min, 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 6 M HCl at 100 °C to remove proteins and carbohydrates, then centrifuged at $5000 \times g$ for 10 min. The precipitate was washed with ethanol and chloroform and then dried at 50 °C overnight. Identity was confirmed via Direct Analysis in Real Time Mass Spectrometry on Dart-AccuTOF mass spectrometer and ^1H NMR on a Varian Mercury 500 mHz spectrophotomer. HRMS-DART $[M + H]^+$ calculated for $C_{10}H_6O_5$: 206.02000; found: 206.16855. ^1H NMR had peaks consistent agreement with literature report[76], however suggested a level of purity that was only appropriate for assessing relative concentration rather than absolute concentration.

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 ($5000 \times g$, 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 $12,000 \times g$ for 10 min 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 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[68]. 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 $1 \times$ 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 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 10 s for 20 HR at 30 °C. A_{340} 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.

SDS-PAGE and Western Blot analysis

In vivo expression of RppA was conducted using *E. coli* BL21 Star(DE3) as a host strain harboring plasmid DNA of four RppA coding sequences in pJ23101 promoter. Cultures were grown in LB broth (Miller) supplemented with appropriate antibiotics (kanamycin at 50 µg/mL). 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. 50 mL 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_{600} of 0.6–0.8 was reached. The temperature was then lowered to 16 °C and incubated for ~ 20 h at 16 °C prior to harvest by centrifugation (4000 × g, 15 min at 4 °C). Cell pellets were resuspended in a wash buffer (100 mM Tris/HCl pH 8.0, 150 mM NaCl, 1 mM EDTA) and lysed by sonication (3 × 30 min on, 1 min off). After sonication, the lysate was clarified via centrifugation (9000 × g, 30 min, 4 °C).

5 µL of RppA lysate which was taken from the combination of 3 triplicate lysates, was denatured with 5 µL 2 × Laemmli sample buffer (BioRad #1610737). After boiling at 98 °C for 10 min, 5 µL of the denatured protein was loaded into a pre-cast 4–20% gel ordered from ThermoFisher (XP04205BOX). The gel was run for 10 min at 80 V followed by 60 min at 220 V.

For protein expressed cell-free, triplicate reactions were pooled into a microcentrifuge tube after reaction time (12 h for pT7 promoter, 38 h for pTet promoter and pBAD promoter, and 20 h for pJ23101 promoter). 5 µL from the reaction was added to 5 µL 2 × Laemmli sample buffer (BioRad #1610737) in a PCR tube. After boiling at 98 °C for 10 min, 8 µL of the denatured protein was loaded into a pre-cast 4–20% gel ordered from ThermoFisher (XP04205BOX). The gel was run for 10 min at 80 V followed by 60 min at 220 V.

For Western Blot analysis, a ThermoFisher Mini gel tank and Mini Blot Module were used to transfer bands onto a nitrocellulose membrane (20 V–60 min). The membrane was blocked using 20 mL PBS-blocking buffer (PBS buffer with 3% BSA and 0.5% v/v Tween 20) for 1 h. The membrane was then washed 3 times with 20 mL PBS-Tween buffer(PBS buffer with 0.1% v/v Tween 20) for 5 min, room temperature, and gentle shake. Next, the membrane was incubate for 10 min in 10 mL PBS-Tween buffer with 10 µL Biotin Blocking Buffer (iba 2-0205-050) before 60 min incubation with the addition of 2.5 µL *Strep*-Tactin horse radish peroxidase conjugate (Bio-Rad 161381). The membrane was washed twice with PBS-Tween buffer for 1 min and then washed with PBS buffer for 1 min before transferred in 20 mL PBS buffer with 200 µL chloronaphtol solution and 20 µL H_2O_2 solution. The chromogenic reaction was observed after ~ 10 min.

Data availability
Data is provided within the manuscript and supplementary information files.

References
1. Bérdy, J. Bioactive microbial metabolites. *J. Antibiot.* **58**, 1–26 (2005).
2. Baltz, R. H. Gifted microbes for genome mining and natural product discovery. *J. Ind. Microbiol. Biotechnol.* **44**, 573–588 (2017).
3. Palazzotto, E., Tong, Y., Lee, S. Y. & Weber, T. Synthetic biology and metabolic engineering of actinomycetes for natural product discovery. *Biotechnol. Adv.* **37**, 107366 (2019).
4. Musiol-Kroll, E. M., Tocchetti, A., Sosio, M. & Stegmann, E. Challenges and advances in genetic manipulation of filamentous actinomycetes: The remarkable producers of specialized metabolites. *Nat. Prod. Rep.* **36**, 1351–1369 (2019).
5. Drufva, E. E., Sword, T. T. & Bailey, C. B. Metabolic engineering of actinomycetes for natural product discovery. In *Natural Products from Actinomycetes: Diversity, Ecology and Drug Discovery* (eds Rai, R. V. & Bai, J. A.) 267–307 (Springer, 2022).
6. Stevens, D. C., Hari, T. P. A. & Boddy, C. N. The role of transcription in heterologous expression of polyketides in bacterial hosts. *Nat. Prod. Rep.* **30**, 1391–1411 (2013).
7. Wagner, L., Jules, M. & Borkowski, O. What remains from living cells in bacterial lysate-based cell-free systems. *Comput. Struct. Biotechnol. J.* **21**, 3173–3182 (2023).
8. Tuckey, C., Asahara, H., Zhou, Y. & Chong, S. Protein synthesis using a reconstituted cell-free system. *Curr. Protoc. Mol. Biol.* **108**, 16.31.1-16.31.22 (2014).
9. Dinglasan, J. L. N. & Doktycz, M. J. Rewiring cell-free metabolic flux in *E. coli* lysates using a block-push-pull approach. *Synth. Biol.* https://doi.org/10.1093/synbio/ysad007 (2023).
10. Garcia, D. C. *et al.* A lysate proteome engineering strategy for enhancing cell-free metabolite production. *Metab. Eng. Commun.* **12**, e00162 (2021).
11. Dinglasan, J. L. N., Reeves, D. T., Hettich, R. L. & Doktycz, M. J. Liquid chromatography coupled to refractive index or mass spectrometric detection for metabolite profiling in lysate-based cell-free systems. *J. Vis. Exp.* https://doi.org/10.3791/62852 (2021).
12. Mouncey, N. J., Otani, H., Udwary, D. & Yoshikuni, Y. New voyages to explore the natural product galaxy. *J. Ind. Microbiol. Biotechnol.* **46**, 273–279 (2019).
13. Bogart, J. W. *et al.* Cell-free exploration of the natural product chemical space. *ChemBioChem* **22**, 84–91 (2021).
14. 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.* **67**, 102142 (2022).
15. Garenne, D. *et al.* Cell-free gene expression. *Nat. Rev. Methods Primers* **1**, 49 (2021).
16. Moore, S. J. *et al.* Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria. *Proc. Natl. Acad. Sci. USA* **115**, E4340–E4349 (2018).
17. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).
18. Lentini, R. *et al.* Fluorescent proteins and in vitro genetic organization for cell-free synthetic biology. *ACS Synth. Biol.* **2**, 482–489 (2013).
19. Jew, K. *et al.* Characterizing and improving pET vectors for cell-free expression. *Front. Bioeng. Biotechnol.* **10**, 895069 (2022).
20. Burrington, L. R., Watts, K. R. & Oza, J. P. Characterizing and improving reaction times for *E. coli*-based cell-free protein synthesis. *ACS Synth. Biol.* **10**, 1821–1829 (2021).
21. Schmidt, M. *et al.* Maximizing heterologous expression of engineered type I polyketide synthases: Investigating codon optimization strategies. *ACS Synth. Biol.* **12**, 3366–3380 (2023).
22. Funa, N. *et al.* A new pathway for polyketide synthesis in microorganisms. *Nature* **400**, 897–899 (1999).
23. 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.* **277**, 4628–4635 (2002).
24. Yang, D. *et al.* Repurposing type III polyketide synthase as a malonyl-CoA biosensor for metabolic engineering in bacteria. *Proc. Natl. Acad. Sci. USA* **115**, 9835–9844 (2018).
25. Katsuyama, Y. & Ohnishi, Y. Type III polyketide synthases in microorganisms. *Methods Enzymol.* **515**, 359–377 (2012).
26. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).

27. Lee, S. K., Newman, J. D. & Keasling, J. D. 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 (2005).
28. Lee, T. S. *et al.* BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J. Biol. Eng.* **5**, 12 (2011).
29. Dubendorff, J. W. & Studier, F. W. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* **219**, 45–59 (1991).
30. William Studier, F., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. [6] Use of T7 RNA polymerase to direct expression of cloned genes. *Gene Expr. Technol.* **185**, 60–89 (1990).
31. Krefft, D., Papkov, A., Zylicz-Stachula, A. & Skowron, P. M. 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 (2017).
32. Walsh, I. M., Bowman, M. A., Soto Santarriaga, I. F., Rodriguez, A. & Clark, P. L. Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. *Proc. Natl. Acad. Sci. USA* **117**, 3528–3534 (2020).
33. Sword, T. T. *et al.* 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 (2023).
34. Chaney, J. L. *et al.* Widespread position-specific conservation of synonymous rare codons within coding sequences. *PLoS Comput. Biol.* **13**, e1005531 (2017).
35. Welch, M. *et al.* Design parameters to control synthetic gene expression in *Escherichia coli*. *PLoS ONE* **4**, e7002 (2009).
36. Mellitzer, A., Weis, R., Glieder, A. & Flicker, K. Expression of lignocellulolytic enzymes in *Pichia pastoris*. *Microb. Cell Fact.* **11**, 61 (2012).
37. Kodumal, S. J. *et al.* Total synthesis of long DNA sequences: Synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. USA* **101**, 15573–15578 (2004).
38. 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.* **26**, 895–901 (2010).
39. Marlatt, N. M., Spratt, D. E. & Shaw, G. S. Codon optimization for enhanced *Escherichia coli* expression of human S100A11 and S100A1 proteins. *Protein Expr. Purif.* **73**, 58–64 (2010).
40. Villalobos, A., Ness, J. E., Gustafsson, C., Minshull, J. & Govindarajan, S. Gene designer: A synthetic biology tool for constructing artificial DNA segments. *BMC Bioinform.* **7**, 285 (2006).
41. Richardson, S. M., Wheelan, S. J., Yarrington, R. M. & Boeke, J. D. GeneDesign: Rapid, automated design of multikilobase synthetic genes. *Genome Res.* **16**, 550–556 (2006).
42. Mignon, C. *et al.* Codon harmonization: Going beyond the speed limit for protein expression. *FEBS Lett.* **592**, 1554–1564 (2018).
43. Wright, G. *et al.* CHARMING: Harmonizing synonymous codon usage to replicate a desired codon usage pattern. *Protein Sci.* **31**, 221–231 (2022).
44. Zhou, M. *et al.* Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* **495**, 111–115 (2013).
45. Angov, E., Hillier, C. J., Kincaid, R. L. & Lyon, J. A. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS ONE* **3**, e2189 (2008).
46. Spencer, P. S., Siller, E., Anderson, J. F. & Barral, J. M. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J. Mol. Biol.* **422**, 328–335 (2012).
47. Shabalina, S. A., Spiridonov, N. A. & Kashina, A. Sounds of silence: Synonymous nucleotides as a key to biological regulation and complexity. *Nucleic Acids Res.* **41**, 2073–2094 (2013).
48. Gilchrist, M. A., Chen, W.-C., Shah, P., Landerer, C. L. & Zaretzki, R. Estimating gene expression and codon-specific translational efficiencies, mutation biases, and selection coefficients from genomic data alone. *Genome Biol. Evol.* **7**, 1559–1579 (2015).
49. Clarke, T. F. & Clark, P. L. Rare codons cluster. *PLoS ONE* **3**, e3412 (2008).
50. Rodriguez, A., Wright, G., Emrich, S. & Clark, P. L. %MinMax: A versatile tool for calculating and comparing synonymous codon usage and its impact on protein folding. *Protein Sci.* **27**, 356–362 (2018).
51. Nakamura, Y., Gojobori, T. & Ikemura, T. Codon usage tabulated from international DNA sequence databases: Status for the year 2000. *Nucleic Acids Res.* **28**, 292 (2000).
52. Cope, A. L. & Gilchrist, M. A. Quantifying shifts in natural selection on codon usage between protein regions: A population genetics approach. *BMC Genom.* **23**, 408 (2022).
53. Incha, M. R. *et al.* Leveraging host metabolism for bisdemethoxycurcumin production in *Pseudomonas putida*. *Metab. Eng. Commun.* **10**, e00119 (2020).
54. Dinglasan, J. L. N., Sword, T. T., Barker, J. W., Doktycz, M. J. & Bailey, C. B. Investigating and optimizing the lysate-based expression of nonribosomal peptide synthetases using a reporter system. *ACS Synth. Biol.* **12**, 1447–1460 (2023).
55. McKevitt, M. *et al.* Systematic cloning of *Treponema pallidum* open reading frames for protein expression and antigen discovery. *Genome Res.* **13**, 1665–1674 (2003).
56. Senda, N. *et al.* Development of an expression-tunable multiple protein synthesis system in cell-free reactions using T7-promoter-variant series. *Synth. Biol. (Oxf.)* **7**, ysac029 (2022).
57. Karim, A. S. *et al.* Modular cell-free expression plasmids to accelerate biological design in cells. *Synth. Biol. (Oxf.)* **5**, ysaa019 (2020).
58. Swartz, J. R., Jewett, M. C. & Woodrow, K. A. Cell-free protein synthesis with prokaryotic combined transcription-translation. *Methods Mol. Biol.* **267**, 169–182 (2004).
59. Sun, Z. Z. *et al.* Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp.* https://doi.org/10.3791/50762 (2013).
60. Garenne, D., Thompson, S., Brisson, A., Khakimzhan, A. & Noireaux, V. The all-*E. coli* TXTL toolbox 3.0: New capabilities of a cell-free synthetic biology platform. *Synth. Biol. (Oxf.)* **6**, ysab017 (2021).
61. Tokmakov, A. A. & Fukami, Y. Activation of T7 RNA polymerase in Xenopus oocytes and cell-free extracts. *Genes Cells* **15**, 1136–1144 (2010).
62. Hagen, A. *et al.* In vitro analysis of carboxyacyl substrate tolerance in the loading and first extension modules of borrelidin polyketide synthase. *Biochemistry* **53**, 5975–5977 (2014).
63. Hagen, A. *et al.* Engineering a polyketide synthase for in vitro production of adipic acid. *ACS Synth. Biol.* **5**, 21–27 (2016).
64. Karig, D. K., Iyer, S., Simpson, M. L. & Doktycz, M. J. Expression optimization and synthetic gene networks in cell-free systems. *Nucleic Acids Res.* **40**, 3763–3774 (2012).
65. Borkowski, O. *et al.* Cell-free prediction of protein expression costs for growing cells. *Nat. Commun.* **9**, 1457 (2018).
66. Brooks, R., Morici, L. & Sandoval, N. Cell free bacteriophage synthesis from engineered strains improves yield. *ACS Synth. Biol.* **12**, 2418–2431 (2023).
67. Guo, S. & Murray, R. M. Construction of incoherent feedforward loop circuits in a cell-free system and in cells. *ACS Synth. Biol.* **8**, 606–610 (2019).
68. Levine, M. Z., Gregorio, N. E., Jewett, M. C., Watts, K. R. & Oza, J. P. *Escherichia coli*-based cell-free protein synthesis: Protocols for a robust, flexible, and accessible platform technology. *J. Vis. Exp.* https://doi.org/10.3791/58882 (2019).
69. Karim, A. S. *et al.* In vitro prototyping and rapid optimization of biosynthetic enzymes for cell design. *Nat. Chem. Biol.* **16**, 912–919 (2020).

70. Vögeli, B. *et al.* Cell-free prototyping enables implementation of optimized reverse β-oxidation pathways in heterotrophic and autotrophic bacteria. *Nat. Commun.* **13**, 3058 (2022).
71. Sivashanmugam, A. *et al.* Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli. Protein Sci.* **18**, 936–948 (2009).
72. Geurink, P. P. *et al.* Profiling DUBs and Ubl-specific proteases with activity-based probes. *Methods Enzymol.* **618**, 357–387 (2019).
73. Khlebnikov, A., Risa, O., Skaug, T., Carrier, T. A. & Keasling, J. D. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J. Bacteriol.* **182**, 7029–7034 (2000).
74. Chappell, J., Jensen, K. & Freemont, P. S. Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. *Nucleic Acids Res.* **41**, 3471–3481 (2013).
75. Wright, G., Rodriguez, A., Clark, P. L. & Emrich, S. A new look at codon usage and protein expression. *Epic Ser. Comput.* **60**, 104–112 (2019).
76. Hillson, N. J., Rosengarten, R. D. & Keasling, J. D. j5 DNA assembly design automation software. *ACS Synth. Biol.* **1**, 14–21 (2012).

Author contributions

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-61376-w.

Correspondence and requests for materials should be addressed to M.J.D. or C.B.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.