Lawrence Berkeley National Laboratory
LBL Publications

Title
Expression of blue pigment synthetase a from Streptomyces lavenduale reveals insights on the effects of refactoring biosynthetic megasynthases for heterologous expression in Escherichia coli.

Permalink
https://escholarship.org/uc/item/54k0q9gt

Authors
Sword, Tien T
Barker, J William
Spradley, Madeline
et al.

Publication Date
2023-10-01

DOI
10.1016/j.pep.2023.106317

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Expression of Blue Pigment Synthetase A from *Streptomyces lavenduale* Reveals Insights on the Effects of Refactoring Biosynthetic Megasynthases for Heterologous Expression in *Escherichia coli.*

Tien T. Sword,[a] J. William Barker,[a] Madeline Spradley, [b] Yan Chen,[c,d] Christopher J. Petzold,[c,d] and Constance B. Bailey*[a]

[a] Tien T. Sword, J. William Barker, and Constance B. Bailey*
Department of Chemistry
University of Tennessee-Knoxville
Knoxville, TN USA
E-mail: cbaile53@utk.edu
[b] Madeline Spradley
Department of Biochemistry, Cellular, and Molecular Biology
University of Tennessee-Knoxville
Knoxville, TN USA
[c] Christopher J. Petzold, Yan Chen
Biological and Systems Engineering Division
Lawrence Berkeley National Laboratory, Berkeley, CA USA
[d] Joint BioEnergy Institute, Emeryville CA USA

Highlights

- Improving the expression of multidomain biosynthetic proteins such as non-ribosomal peptide synthetases (NRPSs) from high GC organisms in *E. coli* often remains a trial-and-error process. To probe this, we assessed a model NRPS, blue pigment synthetase A (BpsA) from *Streptomyces lavenduale* evaluating constructs using native codons versus *E. coli* codon optimization with and without a commercially available tRNA complementation plasmid in a commonly used pET vector.
- Differences in inclusion body formation were detectable when comparing expression conditions, however, for the protein that remained in solution, it was identically stable, and all conditions were entirely post-translationally modified.
- Differences in indigoidine pigment titer catalyzed by BpsA were correlated to differences in soluble protein production, and not misfolding of protein in the soluble fraction due to differences in co-translational folding.

Abstract: High GC bacteria from the genus *Streptomyces* harbor expansive secondary metabolism. The expression of biosynthetic proteins and the characterization and identification of biological "parts" for synthetic biology purposes from such pathways are of interest. However, the high GC content of proteins from actinomycetes in addition to the large size and multi-domain architecture of many biosynthetic proteins (such as non-ribosomal peptide synthetases; NRPSs, and polyketide synthases; PKSs often called "megasynthases") often presents issues with full-length translation and folding. Here we evaluate a non-ribosomal peptide synthetase (NRPS) from *Streptomyces lavenduale*, a multidomain "megasynthase" gene that comes from a high GC (72.5%) genome. While a preliminary step in revealing differences, to our knowledge this presents the first head-to-head comparison of codon-optimized sequences versus a native sequence of proteins of streptomycete origin heterologously expressed in *E. coli*. We found that any disruption in co-translational folding from codon mismatch that reduces the titer of indigoidine is explainable via the formation of more inclusion bodies as opposed to compromising folding or posttranslational modification in the soluble fraction. This result

supports that one could apply any refactoring strategies that improve soluble expression in *E. coli* without concern that the protein that reaches the soluble fraction is differentially folded.

Keywords: heterologous expression • *Streptomyces* • natural products • synthetic biology • refactoring • codon optimization

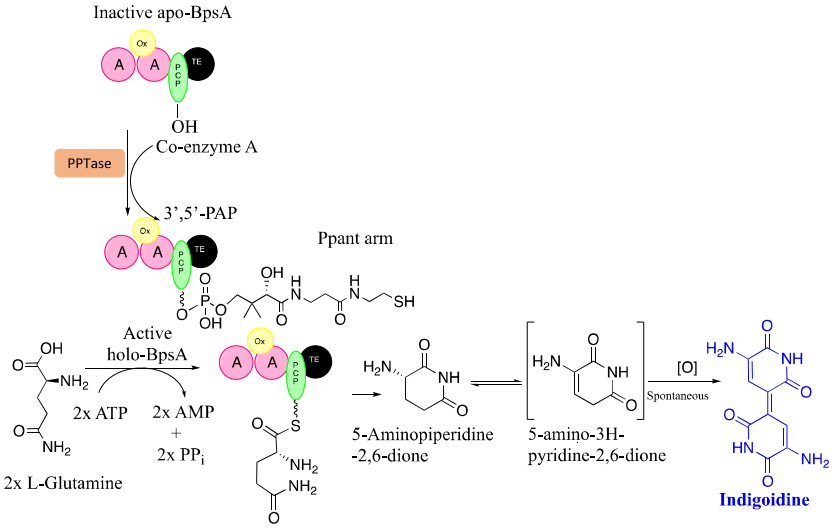
1. Introduction

Bacteria from the order Actinomycetales, especially those from the genus *Streptomyces* are some of the most prolific producers of bioactive natural products. Indeed, over half of the clinically used antibiotics arise from *Streptomyces* and related Actinomycetes [1,2]. Many of these natural products come from classes such as polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs) which generate the polyketide and non-ribosomal peptide metabolites, respectively. PKSs and NRPSs are megadaltons in size (often termed "megasynthases") [3–5]. The large size of PKSs and NRPSs combined with their high GC genomes (~75% GC) often results in poorly understood issues with protein expression and folding in heterologous hosts, especially those that are evolutionarily distinct and more moderate in GC content [6]. While numerous investigations have focused on the improvement of metabolic flux [7–10], the availability of phosphopantetheinyltransferases (PPtases) to provide sufficient posttranslational modification [11], and genome minimization [12–14] to remove potential drains on metabolic resources, such as host optimization efforts will be fundamentally limited if the key bottleneck remains to create full length, well folded and catalytically active protein. *Streptomycetes* harbor biosynthetic genes involved in secondary metabolism in general, and particularly PKSs and NRPSs have had extensive investigation with the goal of creating engineered products including medicinal agents [7,9] as well as commodity and specialty chemicals [15–17]. None of these synthetic biology applications are feasible without the critical step of biochemical characterization of biosynthetic proteins.

Well-established heterologous expression systems exist for various *Streptomyces* sp. [18–21], however, they typically remain a tool of last resort for the application of production of adequate protein for overexpression and biochemical characterization, which is a fundamental step for characterizing synthetic biology parts. This is due to their slow growth curves, less than ideal growth properties (e.g. mycelial clumping), and typical requirement of transfer of genetic material through specialized techniques (e.g. intergenic conjugation or lengthy protoplast preparations) results in less than ideal tractability for isolation and purification of protein [18]. Despite the vast evolutionary differences between *E. coli* and *Streptomyces* and related actinomycetes, the unparalleled genetic tractability, as well as the rapid doubling time and facile growth conditions, still fuels motivation for expressing high GC and large proteins within *E. coli*. While *Streptomyces* species might be all-around better hosts for metabolite production in many settings [22], the downsides of expression for protein overexpression still typically weigh the scale towards optimizing expression in *E. coli* for this key application. With *E. coli*, due to the large differences in GC content between *Streptomyces* sp. and *E. coli*, this means that it is not always obvious whether an expression is improved by maintaining the native coding sequence or by performing codon optimization to better match the codon usage of *E. coli* when considering the need highly expressed appropriately folded soluble protein. Additionally, the native coding sequence may be intractable to generate as a synthetic gene due to gene synthesis companies'

lack of capacity to produce high GC DNA sequences, necessitating a synthetic gene that uses synonymous codons that are more synthetic accessible if the organism is not readily cultured and/or the gene is challenging to clone from genomic DNA.

Despite the mechanistic knowledge that codon usage has an evolutionary basis to regulate the rate of translation and co-translational folding [23–26], there are contradictory anecdotal reports of success and failure of codon-optimized constructs among members of the natural products enzymology and synthetic biology communities. Even though the genetic code is universal, the usage frequencies of synonymous codons can differ substantially between organisms. Codon bias is when there is a preferred usage of one codon over another for the same amino acid. When codon usage differs in a heterologous host, codons that are rarely used in *E. coli* such as CUA (Leucine); AGA, AGG (Arginine); AUA (Isoleucine); CCC, (Proline); GGA, GGG (Glycine) essentially regulates the expression of different endogenous protein (AGG, CCC, GGA, and GGG are common in *Streptomyces sp.)*.[27,28] The level of reduction of protein expression in *E. coli* corresponds with the relative positions of these rare codons in genes. Due to the lack of cognate tRNAs corresponding with rare codons in the host organism, the expression of heterologous proteins is limited. With such extreme GC bias, deciding whether to codon optimize or not is not always clear to the researcher. While some head-to-head comparisons of *E. coli* codon-optimized constructs compared to natively coded constructs exist for eukaryotic proteins [27], few analogous experiments have not been performed for high GC prokaryotes. To date, the only comparison of codon-optimized *Streptomyces* constructs to a natively coded construct has been for a beta-glucanase enzyme which indicated small but measurable differences in protein folding that affected protein activity between synonymous coding sequences[29]. Because there is high interest in expressing biosynthetic genes for synthetic biology purposes from such metabolically gifted bacteria, we sought to do a head-to-head comparison of a model megasynthase with the supplementation of tRNA for rare codon in *E. coli*. For our model, we chose blue pigment synthetase A (BpsA) from *S. lavendulae* which produces the blue pigment, indigoidine [30]. BpsA is an example of type 1 non-ribosomal peptide synthetase (NRPS) [31], which is a large, multi-domain protein with flexible linkers and an example of a protein that provides unique challenges to heterologous expression among relevant biosynthetic proteins from *Streptomyces* that we sought to interrogate (**Scheme 1**).



Scheme 1. Production of indigoidine by the single-module NRPS BpsA. Schematic diagram showing apo-BpsA activated to holo-BpsA via attachment of a phosphopantetheinyl prosthetic group derived from Co-enzyme A, mediated by a phospopantetheinyltransferase. Two molecules of L-glutamine are converted by holo-BpsA into the easily detectable blue pigment indigoidine. BpsA consists of an adenylation (A) domain with an oxidase (Ox) domain present between subdomains of the A-domain, a peptidyl carrier protein (PCP) domain, and a thioesterase (TE) domain.

2. Materials and Methods

2.1 Bacterial strain, plasmid vectors, and chemicals

Chemically competent *E. coli* DH5α cells, *E. Coli* BL21(DE3) Rosetta were purchased from Novagen, USA. BL21(DE3) Rosetta strain contains a plasmid harboring tRNA genes for the following rare codons: AGG, AGA, AUA, CUA, CCC, and GGA on a chloramphenicol-resistant plasmid. Chemically competent *E. coli* BAP1 was obtained from Prof. Christopher Boddy (University of Ottawa, Ottawa, Canada). *E. coli* BAP1 essentially is a derivative of the commonly used of *E. coli* BL21(DE3) strain, which has the promiscuous PPTase *sfp* to provide the required phosphopantetheinyl post-translational modification.[11] Hence, with present of *sfp* allows BAP1 to generate phosphopantetheinylated active *holo*-ACPs and PCPs. Native *bpsA* in the pET28 vector was obtained from Prof. Michael Burkart (University of California San Diego) [32]. The *E. coli* codon-optimized sequences of *bpsA* were ordered as two gene blocks (part 1 and part 2) from Integrated DNA Technology's (USA) (supplementary information). GoldenGate cloning kit, DNA ladder (1kb), and protein ladder were purchased from New England BioLabs, USA. PrimeSTAR GXL Premix was purchased from Takara Bio Inc., USA. PCR cleanup kits, miniprep kits, and gel recovery kits were purchased from Zymo Research, USA. Primers were obtained from Thermofisher, USA. Chemicals required for SDS-PAGE, purification, Luria Bertani (LB) broth, and agar were purchased from Thermofisher. Pre-made protein gel was purchased from NuSep, USA. All constructs were confirmed by Sanger sequencing.

2.2 Codon optimization and gene synthesis

The DNA sequence of the native *bpsA* in the pET28 vector has been previously described [32]. The *E. coli* codon-optimized sequence of *bpsA* was designed using Integrated DNA Technology's (IDT, USA) codon optimization tool.

2.3 Plasmid Construction

Natively coded *bpsA* cloned into pET28 was generously provided by Michael Burkart (University of California San Diego, See supplementary) [32]. The *E. coli* codon-optimized *bpsA* and pET28 vector were amplified by PCR. The construct was cloned using GoldenGate cloning kit (NEB, USA) and included an N-terminal His_6-tagged and a stop codon (TAA). Primers were designed with the J5 algorithm [33]. Primers used in this study are listed in **Table 1**.

Table 1 Primers used in this study.

Primer name	Sequence
E. coli bpsA part 1	CACACCAGGTCTCAGCATGACGTTGCAAGAGACTTCGG

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

2.4 Expression and Purification of BpsA

Expression of BpsA was conducted using either *E. coli* BAP1[11] as a host strain or *E. coli* BAP1 transformed with the tRNA complementation plasmid isolated from *E. coli* BL21(DE3) Rosetta (Novagen, USA). Cultures were grown in LB broth (Miller) supplemented with appropriate antibiotics (kanamycin at 50 µg/mL and 25 µg/mL chloramphenicol as appropriate). Overnight seed cultures were grown in 25 mL LB broth and kanamycin (50 µg/mL) inoculated with a single colony at 37 °C, shaking at 210 rev/min. 2L of expression cultures were inoculated from these cultures in a ratio of 1:100 and incubated at 37 °C shaking at 210 rev/min until an OD_{800} of 0.3-0.4 was reached. Note that OD_{800} rather than the more typical OD_{600} was used to monitor growth to ensure that interference from the signal from indigoidine would not interfere with measurements of cellular turbidity. This strategy was adapted from Beer and coworkers [34] and a reproduction of the correlation between the OD_{600} and OD_{800} is shown in (**Fig. 1**). The temperature was then lowered to 16 °C and cooled for 15 minutes prior to induction by the addition of IPTG to a final concentration of 0.5 mM. The cultures were then incubated for ~20 hours at 16 °C prior to harvesting by centrifugation (4000 x g, 45 min at 4 °C).

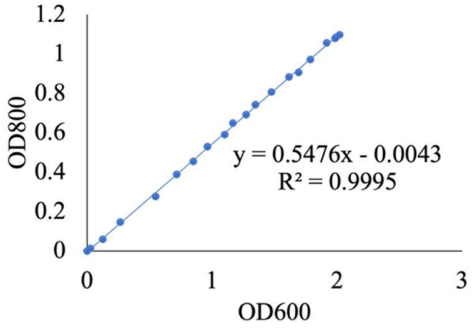


Fig. 1. Correlation of the optical density of BAP1 between OD_{800} and OD_{600} measurements. The accuracy of the R^2 shows that we can measure the growth in this study at OD_{800}.

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

(CV), 60-100, 1 CV, and hold at 100% for 3 CV. After purification, the collected fractions were dialyzed (10K MWCO dialysis tubing Thermofisher) for 2 passes for a minimum of six hours into storage buffer (50 mM sodium phosphate pH 7.8, 100 mM NaCl, 10% v/v glycerol). The protein was then concentrated using a 100 K MWCO concentrator (Thermofisher, USA). Purity was assessed via SDS-PAGE using 4-20% pre-made gels (NuSep, USA).

2.5 Indigioidine Purification and Preparation of the Indigoidine Standard Curve

A 10 mL seed culture containing BAP1 harboring *E. coli* coded *bpsA* was grown from a fresh single colony overnight (37 °C, 2010 rev/min) in LB medium supplemented with 50 µg/mL kanamycin. After ~20 hours, 1L of culture was inoculated with the overnight seed culture in a ratio of 1:100 and incubated at 37 °C shaking at 210 rev/min. At an OD_{800} of 0.3-0.4, the overnight culture was induced with 0.5 mM IPTG, grown at 16 °C for 20 h, and then lyophilized for 36 hours. The dry cell mass was then washed with three rounds of water, methanol, ethanol, isopropanol, and hexanes to remove metabolites, salts, and proteins. Finally, the product was dried for ~one week under vacuum. Purity was verified via ^{1}H NMR: (400 mHz, D_6MSO) δ 11.30 ppm (s, NH), 8.18ppm (s, CH), and 6.46ppm (s, NH_2) which is in agreement with the previous literature report [35,36]. Afterward, 0.5 mg of dry indigoidine was dissolved in 1 mL DMSO. This solution was then serially diluted to six different concentrations (0.01, 0.025, 0.5, 0.1, 0.2, and 0.25 mg/mL). 200 µL of the solution was added to a 96-well plate in triplicate to measure A_{595} to generate a standard curve.

2.6 Measurement of Titer

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

2.7 Thermal shift assay:

A Thermal shift assay was used to determine the melting temperature (T_m) of BpsA using SYPRO orange [37]. Each BpsA expressed under each expression condition was purified to homogeneity. Triplicate reactions of three separate preparations were established for the assay. In a 96-well plate, reaction conditions were as follows 20µL storage buffer (50 mM phosphate pH7.8, 100mM NaCl, 10%(v/v) glycerol), 5 µL 50x SYPRO orange dye, 15 µL Millipore water, and 10 µL of protein with the concentration between 0.5-7 mg/mL. Lysosome (1mg/ml) was used as the positive control. Briefly, each well was measured from 20 °C to 98 °C with the rate

was 1 °C. The corresponding of the lowest measurement of the derivative data used to determine T_m of BpsA (supplementary information).

2.8 Proteomics Analysis:

50 ug of each purified protein sample was precipitated using the established acetone precipitation method [38]. Proteins were resuspended in 100 mM ammonium bicarbonate buffer supplemented with 20% methanol, followed by reduction using 5 mM tris 2-(carboxyethyl) phosphine (TCEP) for 30 min at room temperature, and alkylation with 10 mM iodoacetamide (IAM; final concentration) for 30 min at room temperature in the dark. Overnight digestion with trypsin was accomplished with a 1:50 trypsin: total protein ratio. The resulting peptide samples were analyzed on an Agilent 1290 UHPLC system coupled to a Thermo scientific Obitrap Exploris 480 mass spectrometer for the discovery of proteomics [39]. Briefly, 20 µg of tryptic peptides were loaded onto an Ascentis® (Sigma–Aldrich) ES-C18 column (2.1 mm × 100 mm, 2.7 µm particle size, operating at 60°C) and were eluted from the column by using a 10-minute gradient from 98% buffer A (0.1 % FA in H_2O) and 2% buffer B (0.1% FA in acetonitrile) to 65% buffer A and 35% buffer B. The eluting peptides were introduced to the mass spectrometer operating in positive-ion mode. Full MS survey scans were acquired in the range of 300-1200 m/z at 60,000 resolutions. The automatic gain control (AGC) target was set at 3e6, and the maximum injection time was set to 60 ms. The top 10 multiply charged precursor ions (2-5) were isolated for higher-energy collisional dissociation (HCD) MS/MS using a 1.6 m/z isolation window and were accumulated until they either reached an AGC target value of 1e5 or a maximum injection time of 50 ms. MS/MS data were generated with a normalized collision energy (NCE) of 30, at a resolution of 15,000. Upon fragmentation precursor ions were dynamically excluded for 10 s after the first fragmentation event. The acquired LCMS raw data were converted to mgf files and searched against the latest UniProt *E. coli* protein database supplemented with BpsA and other common contaminant protein fasta sequences using Mascot search engine version 2.3.02 (Matrix Science). Phosphopantetheine modification to L-serine was defined as a variable modification in addition to other common structural modification parameters, such as carbamidomethyl and oxidized methionine. The resulting search results were filtered and analyzed by Scaffold v 5.0 (Proteome Software Inc.). The quantitative report of phosphopantetheine-modified peptide was analyzed by Skyline v 22.2 (University of Washington).

Data Availability

The generated mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[40] partner repository with the dataset identifier PXD040627 and 10.6019/PXD040627.

3. Results and Discussion

3.1 Expression and purification of BpsA

To test the hypothesis that the difference in coding sequence might impact BpsA stability via disruptions to co-translational folding, we cloned two synonymously coded constructs of *bpsA* into the commonly used expression vector, pET28a. Because of the T7 promoter, high copy

number origin, and lac-inducible operon, this is typically a commonly used vector for overexpression and purification for biochemical characterization and is usually the first choice for the application of biochemical characterization. First, we used a coding sequence that was cloned directly from genomic DNA[32] of *S. lavendulae* as the native coding sequence. Next, we ordered *bpsA* as a gene block from Integrated DNA Technologies using their Codon Optimization Tool to create a synonymously coded *E. coli* codon-optimized construct of *bpsA*. The natively coded *bpsA* gene had a GC content of 68% whereas the *E. coli* codon-optimized gene had a GC content of 51%. The overall genome of *Streptomyces lavenduale* has a GC content of 72.5%, thus the GC content of *bpsA* is lower than a typical gene from *Streptomyces lavenduale.*

We measured purified protein as a proxy for how much fully translated, soluble protein was present. The electrophoretic analysis of the complete purification process as well as the purity was assessed via SDS-PAGE (**Fig. 2, Fig. S1**). The expression of both plasmids was performed in *E. coli* BAP1, a derivative of the commonly used BL21(DE3) strain specifically designed for PKS and NRPS expression a copy of the promiscuous phosphopantetheinyl transferase, *sfp* is integrated into its genome (PKS and NRPS expression require phosphopantetheinylation for catalytic activity). We saw a lower purified yield of protein with the native codon construct than with the *E. coli* codon-optimized construct (1.4±0.1 mg/L versus 2.1±0.17 mg/L respectively, **Table 2**).

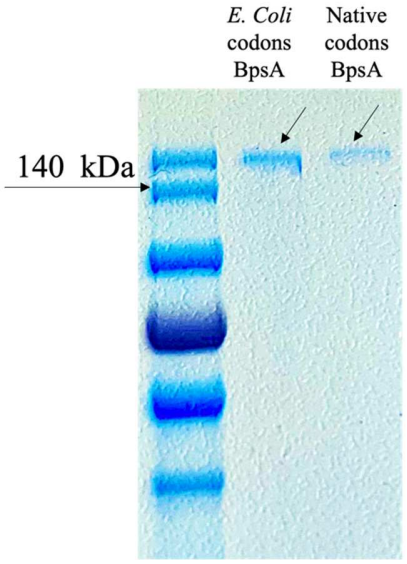


Fig. 2. Two constructs of BpsA were purified to homogeneity with a molecular mass of 145 kDa. The expression of *E. coli* codon-optimized *bpsA* and native-coded *bpsA* was performed in *E. coli* BAP1, which is a subsequence strain of the commonly used BL21(DE3). Because PKS and NRPS expression requires phosphopantetheinylation for catalytic activity, *sfp* is integrated into BL21(DE3) genome to create BAP1 strain for the expression of PKSs and NRPSs. This figure indicates the purity of each construct purified to homogeneity.

Table 2. T_m values, purified protein yields, and indigoidine titers for each expression condition of BpsA

Construct	**T_m (˚C)**	**Purified Protein Yield**	**Titer of indigoidine**

		(mg/L)	(mg/ml)
E. coli codon-optimized *bpsA*	41.3 ± 0.06	2.1 ± 0.17	0.41 ± 0.03
Natively coded *bpsA*	41.3 ± 0.12	1.4 ± 0.10	0.21 ± 0.06
E. coli codon optimized *bpsA* expressed with the Rosetta plasmid	41.2 ± 0.10	4.8 ± 0.48	0.49 ± 0.005
Natively coded *bpsA* expressed with the Rosetta plasmid	41.1 ± 0.05	5.2 ± 0.38	0.49 ± 0.003

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

Because the *E. coli* codon-optimized construct showed a higher yield of purified protein than the native codon construct, we decided to pursue another common strategy to improve the expression of genes with a codon usage that differs from *E. coli*, using tRNA complementation plasmids. The commercially available BL21(DE3) Rosetta strain contains a plasmid harboring tRNA genes for the following rare codons: AGG, AGA, AUA, CUA, CCC, and GGA on a chloramphenicol-resistant plasmid. To ensure that we retained phosphopantetheinylation, we isolated this plasmid from the commercially available BL21(DE3) Rosetta strain and transformed it into *E. coli* BAP1. When supplemented with the rare tRNA complementation plasmid, the yield of purified protein becomes comparable within prep-to-prep variability, with the native coding sequence slightly higher yielding (5.2±0.38 mg/L versus 4.8±0.48 mg/L for native coding and E. *coli* codon-optimized constructs, respectively) (**Table 2**). However, the levels of insoluble protein become more comparable between the two conditions when qualitatively comparing via SDS page gel evaluating the soluble versus insoluble fractions (**Fig. 3**). The additional tRNA not only improves the yield of the native construct but also improves the yield of the *E. coli*-coded construct. This can be explained that the Rosetta plasmid provides a rich source of tRNA in the translation process, hence the yields are increased for both constructs. Interestingly, even though there is an increase in the insoluble protein of the E. *coli* codon-optimized constructs when adding the Rosetta plasmid, the yield of purified protein also increased. This means that the total proteins that are expressed are significantly increased along with the inclusion bodies, and the increase in recovered soluble protein is significant.

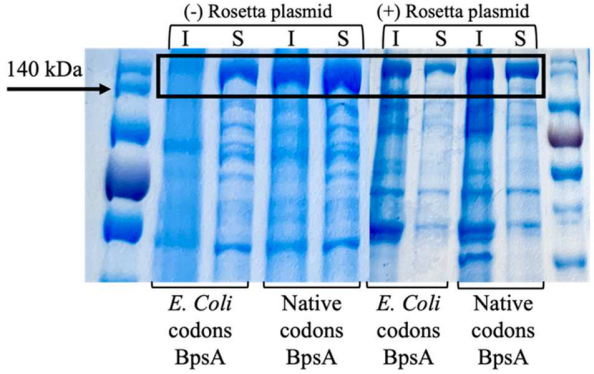


Fig. 3. Solubility of BpsA. BpsA expressed in *E. coli* BAP1 with and without the addition of tRNA from BL21(DE3) Rosetta. Without the Rosetta plasmid, the native coding sequence is less soluble than the *E. coli* codon-optimized construct. With the Rosetta plasmid, the expression between the two constructs is improved.

3.3 Measurement of Indigoidine Titer

As an output for the functional activity of BpsA in a context relevant to metabolite production, we measured the titer of indigoidine to better understand the functional consequences of codon optimization. A standard curve for purified indigoidine was established (**Fig. S3**). We saw a distinct difference in titer with approximately 2-fold more indigoidine production in the *E. coli* codon-optimized construct than in the native construct (0.41 ± 0.03 mg/mL and 0.21± 0.06 mg/mL, respectively for the *E. coli* codon-optimized construct versus the native construct). (**Fig. 6, Table 2**). This result also correlates with purified protein yield (**Table 2**). These trends are largely explainable by the fact that more protein was observed in the insoluble fraction as inclusion bodies via qualitative examination of the soluble versus insoluble fractions via SDS-PAGE (**Fig. 3**). When supplemented with the rare tRNA complementation plasmid, the titers became comparable between the two coding constructs (0.49 mg/mL) (**Fig. 6**). Because the levels of insoluble protein become more comparable between the two conditions when supplemented with Rosetta plasmid when qualitatively comparing via SDS page gel evaluating the soluble versus insoluble fractions (**Fig. 3**), which explains why both the titer of iindigoidine, and yield of purified protein were comparable for each coding sequence.

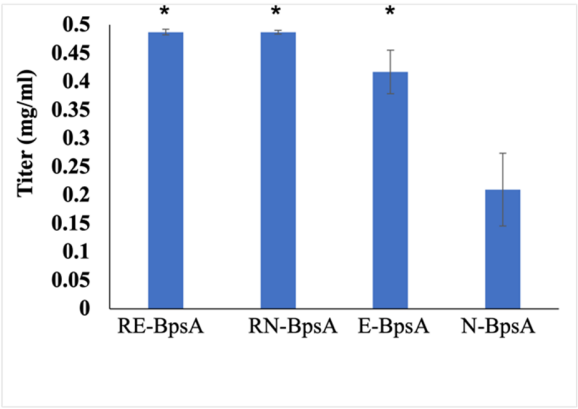


Fig. 4. Titer of indigoidine. Without the Rosetta plasmid, the native coding sequence produced less indigoidine than the *E. coli* codon-optimized construct**.** With the Rosetta plasmid, the production of indigoidine becomes comparable between the two constructs. Evaluated with a Student's t test. *p-value <0.01.

3.4 Thermal Melting Shift Assay

To determine if the soluble, fully translated protein that we were able to purify to homogeneity (**Fig. 2**) was identically folded, we measured the thermal melting point using the SYPRO orange thermal shift assay [41]. All four expression conditions led to a thermal melting point of 41° C (**Table 2**). Because the thermal melting points were identical between the four expression conditions, this suggests that any differences between tRNA pool match and issues with co-translational folding lead to the formation of inclusion bodies, but the protein that is well folded

enough to remain in the soluble fraction is folded appropriately. Consequently, we can conclude that provided that the protein reaches the soluble layer, it is equally well folded regardless of strategy used to promote increased soluble expression. Curiously, this result is different than what was found by Pradeep and coworkers' experiments investigating synonymously coded Beta glucanase enzymes from *Streptomyces althiiotiicus* TBG-MR17 and *Streptomyces cinereoruber* subsp. *Cinereoruber* TBG-AL3 appeared to have minor but measurable differences in thermal stability between natively coded constructs[29]. This difference in results between different classes of proteins from related organisms indicates that more work needs to be done to understand the effects of recoding genes from high GC prokaryotes. It is possible that the large size of BpsA may explain some of these differences, but further investigation needs to be done before we can determine this conclusively.

3.5 Posttranslational modification of BpsA

Phosphopantetheinylation is a post-translational modification process, which is important for activation of NRPSs as the peptidyl carrier protein inactive without a posttranslational modification to a conserved serine residue. [42] Because we observed indigoidine production and similar T_m values among all expression conditions, we did not suspect to observe differences in the degree of phosphpantetheinyl posttranslational medication. However, to confirm, a phosphopantetheinyl ejection assay was performed to determine the ratio of *holo/apo*-protein on the PCP domains. This assay was originally developed by Dorrestein and coworkers and was optimized by the Keasling lab.[43,44] Four proteins were purified via the His-tag purification method described above and 20 µL of the purified sample was used to measure their phosphopantetheinylation ratio. Targeted tandem mass spectrometry methods for the phosphopantetheine ejection assay were generated by the Skyline method.[45] The proteomic data were normalized to "global standards", which is a peptide fragment from BpsA (VELDEISLAIENHDWVR) to minimize differences in the BpsA purity between samples. This fragment is identified in all of the samples at high intensity by using the Skyline 'Global Standards' normalization (**Fig. S3**). The tryptic PCP fragment containing NSL active site motif is the "parent" ion (ENASVQDDFFESGGN**S**LIAVGLVR, serine that is post-translationally modified with phosphopantetheine is bolded). This fragment carries the phosphopantetheine attachment sites including the acyl-phosphopantetheine. The phosphopantetheine fragment and acyl group is the 'daughter' ion (**Fig. 5**).

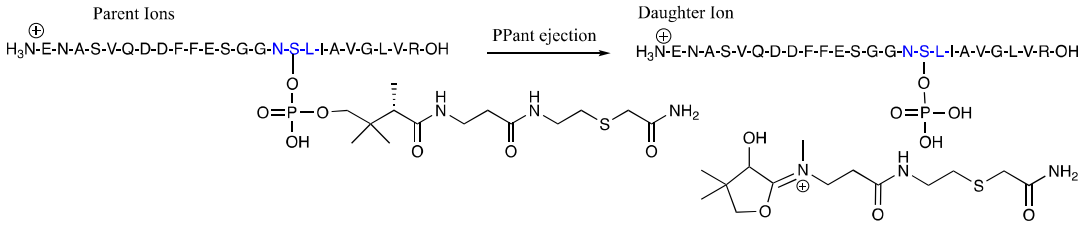


Fig. 5. PPant ejection fragment after treatment with iodoacetamide during proteomic analysis. The "parent" ion is the tryptic PCP fragment containing NSL active site motif (S in NSL is the active serine). The phosphopantetheine fragment and acyl group is the 'daughter' ion.

We found purified BpsA protein across all expression conditions to be present in detectable amounts in the holo-form only without any evidence of detectable apo form, indicating that differences in the percentage of protein that was posttranslationally modified do not explain differences in the titer of indigoidine (**Fig. 6**). Different native and *E. coli* proteins in the LC-MS analysis across the samples were observed, suggesting that the purity of BpsA is lower in the tRNA complementation group. Since the analysis of the proteomic data focuses on the different states of the BpsA protein and it was detected with high sensitivity in all samples, the absolute abundance differences between the samples do not impact the conclusions drawn from the data. The "Expected" bar plot represents the predicted isotope distribution of the Holo-peptide as calculated by the software Skyline and corresponds to an idotp value of 1.0. The idotp values of the Holo-peptide for the sample data indicate how close the observed isotope distribution (measured by a peak intensity of isotope ions) is to the predicted (Expected) distribution. An idotp value greater than 0.9 is commonly accepted as a cutoff to explain the idotp value. Thus, the idotp values of the Holo-peptide for the sample data indicate how close the observed isotope distribution (measured by peak intensity of isotope ions) is to the predicted (Expected) distribution. While we did see variations in the total peak area of the holo-peptides among different purified sample sources, we believe this is explainable by 1) minor variation in the protein quantification assay used to measure the protein concentration of each sample, and 2) variation of the purity of BpsA enriched via the His-tag purification method. We observed different native *E. coli* proteins in the LC-MS analysis across the samples, suggesting that the purity of BpsA may be lower in the tRNA complementation group. Since the analysis of the proteomic data focuses on different states of the BpsA protein and it was detected with high sensitivity in all samples, the absolute abundance differences between samples do not impact conclusions drawn from the data.

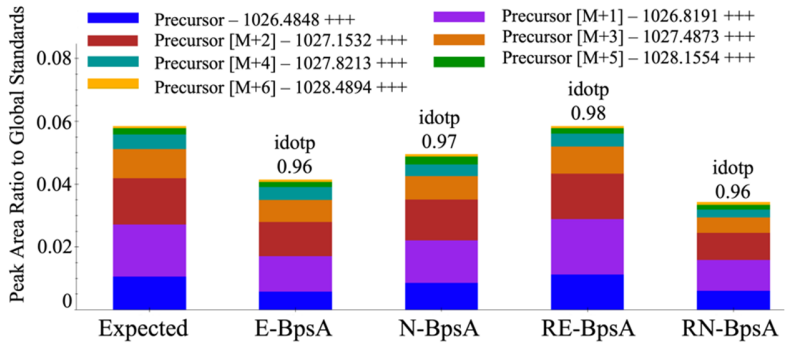


Fig. 6. Holo-BpsA across all expression conditions: E-BpsA is *E. coli* codon-optimized BpsA. N-BpsA is natively coded BpsA. RE-BpsA is *E. coli* codon-optimized BpsA with tRNA from *E. Coli* BL21(DE3) Rosetta complementation. RN-BpsA is natively coded BpsA with tRNA from *E. Coli* BL21(DE3) Rosetta complementation. The "Expected" bar plot represents the predicted isotope distribution of the Holo-peptide as calculated by the software Skyline and corresponds to an idotp value of 1.0.[45] The idotp values of the Holo-peptide for the sample data indicate how close the observed isotope distribution (measured by peak intensity of isotope ions) is to the predicted (Expected) distribution.

4. Conclusion

While there has been much discussion within the literature about how codon usage promotes correct protein folding via evolutionarily tuned co-translational folding events **[24,25,46,47]**, the

details of what the consequences of this are at the protein level are not entirely clear. While one might expect there would be no minor differences in a well-folded protein such as model robust protein such as GFP **[48]**, megasynthases like BpsA have multiple domains held with flexible linkers with multiple opportunities for partial misfolding that may compromise stability and, consequentially, activity. Because many biosynthetic proteins of interest have more complexity than GFP, BpsA served as an appropriate proxy for proteins that are of more interest to investigating and engineering biosynthetic pathways **[49]**.

We found that all expression strategies generated protein with identical T_ms, suggesting no detectable difference in stability in the protein that was folded enough to be recovered from the soluble fraction. The differences in protein yield and metabolite titer could be entirely explained by differences in the solubility of translated BpsA, and not protein stability or phophopantetheinylation. As there is evidence that codon usage affects secondary structure formation in *E. coli*, we wanted to decouple the effects on the solubility expressed protein versus the insoluble protein **[50]**. The identical stability of BpsA purified from different expression strategies has some important ramifications on refactoring in the context of heterologously expressed megasynthases. This means that regardless of the strategy used to improve expression (synonymous coding or complementation of rare tRNAs), provide that *Streptomyces* NRPS proteins can reach the soluble fraction, it appears to be structurally uncompromised. For proteins that are less tractable than BpsA in the aspect of solubility, which means proteins that are still present as inclusion bodies after codon optimization, suggests a more sophisticated approach such as codon harmonization wherein there is an effort to replace rare codons in the host organism with positions of rare codons in the native organism as opposed to just statistically using the most efficient codon for the new heterologous host as occurs in codon optimization **[51,52]**. Codon harmonization strategy matches rare codons to align stalling kinetics between the native and heterologous ribosomes may not affect protein folding in a detectable way as long as it is stable enough to reach the soluble fraction. This has implications for exploring better expression strategies for complex biosynthetic proteins more generally so we can have improved biochemical characterization of parts for synthetic biology from *Streptomyces* and related actinomycetes.

Acknowledgments

Funding was provided by the NIH (R15GM146192), The University of Tennessee-Knoxville, and the University of Tennessee Oak Ridge Innovation-Institute Science Alliance. We thank Prof. C. T. Trinh (University of Tennessee-Knoxville) for assistance in the thermal melt assay and Prof. Michael D. Burkart (University of California, San Diego) for generously providing natively coded BPSA cloned into pET28a and for useful discussion. The use of proteomic resources were supported by the Joint BioEnergy Institute, supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under contract DE-AC02-05CH11231.

Keywords: heterologous expression • *Streptomyces* • natural products • synthetic biology • refactoring

References

[1] J.W.-H. Li, J.C. Vederas, Drug discovery and natural products: end of an era or an endless frontier?, Science. 325 (2009) 161–165. https://doi.org/10.1126/science.1168243.

[2] R.H. Baltz, Gifted microbes for genome mining and natural product discovery., J. Ind. Microbiol. Biotechnol. 44 (2017) 573–588. https://doi.org/10.1007/s10295-016-1815-x.

[3] T. Robbins, Y.-C. Liu, D.E. Cane, C. Khosla, Structure and mechanism of assembly line polyketide synthases., Curr. Opin. Struct. Biol. 41 (2016) 10–18. https://doi.org/10.1016/j.sbi.2016.05.009.

[4] C. Hertweck, Decoding and reprogramming complex polyketide assembly lines: prospects for synthetic biology., Trends Biochem. Sci. 40 (2015) 189–199. https://doi.org/10.1016/j.tibs.2015.02.001.

[5] T.B. Cook, B.F. Pfleger, Leveraging synthetic biology for producing bioactive polyketides and non-ribosomal peptides in bacterial heterologous hosts., Medchemcomm. 10 (2019) 668–681. https://doi.org/10.1039/c9md00055k.

[6] D.C. Stevens, T.P.A. Hari, C.N. Boddy, The role of transcription in heterologous expression of polyketides in bacterial hosts., Nat. Prod. Rep. 30 (2013) 1391–1411. https://doi.org/10.1039/c3np70060g.

[7] C. Olano, F. Lombó, C. Méndez, J.A. Salas, Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering., Metab. Eng. 10 (2008) 281–292. https://doi.org/10.1016/j.ymben.2008.07.001.

[8] S. Murli, J. Kennedy, L.C. Dayem, J.R. Carney, J.T. Kealey, Metabolic engineering of Escherichia coli for improved 6-deoxyerythronolide B production., J. Ind. Microbiol. Biotechnol. 30 (2003) 500–509. https://doi.org/10.1007/s10295-003-0073-x.

[9] E. Palazzotto, Y. Tong, S.Y. Lee, T. Weber, Synthetic biology and metabolic engineering of actinomycetes for natural product discovery., Biotechnol. Adv. 37 (2019) 107366. https://doi.org/10.1016/j.biotechadv.2019.03.005.

[10] L.B. Pickens, Y. Tang, Y.-H. Chooi, Metabolic engineering for the production of natural products., Annu. Rev. Chem. Biomol. Eng. 2 (2011) 211–236. https://doi.org/10.1146/annurev-chembioeng-061010-114209.

[11] B.A. Pfeifer, S.J. Admiraal, H. Gramajo, D.E. Cane, C. Khosla, Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*., Science. 291 (2001) 1790–1792. https://doi.org/10.1126/science.1058092.

[12] H. Ikeda, S. Kazuo, S. Omura, Genome mining of the Streptomyces avermitilis genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters., J. Ind. Microbiol. Biotechnol. 41 (2014) 233–250. https://doi.org/10.1007/s10295-013-1327-x.

[13] M. Komatsu, T. Uchiyama, S. Omura, D.E. Cane, H. Ikeda, Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism., Proc Natl Acad Sci USA. 107 (2010) 2646–2651. https://doi.org/10.1073/pnas.0914833107.

[14] L. Li, X. Liu, W. Jiang, Y. Lu, Recent advances in synthetic biology approaches to optimize production of bioactive natural products in actinobacteria., Front. Microbiol. 10 (2019) 2467. https://doi.org/10.3389/fmicb.2019.02467.

[15] S. Yuzawa, A. Zargar, B. Pang, L. Katz, J.D. Keasling, Commodity chemicals from engineered modular type I polyketide synthases., Meth. Enzymol. 608 (2018) 393–415. https://doi.org/10.1016/bs.mie.2018.04.027.

[16] J.F. Barajas, J.M. Blake-Hedges, C.B. Bailey, S. Curran, J.D. Keasling, Engineered polyketides: Synergy between protein and host level engineering., Synthetic and Systems Biotechnology. 2 (2017) 147–166. https://doi.org/10.1016/j.synbio.2017.08.005.

[17] B. Pang, L.E. Valencia, J. Wang, Y. Wan, R. Lal, A. Zargar, J.D. Keasling, Technical Advances to Accelerate Modular Type I Polyketide Synthase Engineering towards a Retro-biosynthetic Platform, Biotechnol. Bioprocess Eng. 24 (2019) 413–423. https://doi.org/10.1007/s12257-019-0083-9.

[18] D.A. Hopwood, T. Kieser, M. Bibb, K. Chater, Practical Streptomyces Genetics | NHBS Academic & Professional Books, (2000). https://www.nhbs.com/practical-streptomyces-genetics-book (accessed November 17, 2020).

[19] R.H. Baltz, Genetic manipulation of secondary metabolite biosynthesis for improved production in Streptomyces and other actinomycetes., J. Ind. Microbiol. Biotechnol. 43 (2016) 343–370. https://doi.org/10.1007/s10295-015-1682-x.

[20] Y. Zhao, G. Li, Y. Chen, Y. Lu, Challenges and advances in genome editing technologies in streptomyces., Biomolecules. 10 (2020). https://doi.org/10.3390/biom10050734.

[21] N. Lee, S. Hwang, Y. Lee, S. Cho, B. Palsson, B.-K. Cho, Synthetic biology tools for novel secondary metabolite discovery in streptomyces., J. Microbiol. Biotechnol. 29 (2019) 667–686. https://doi.org/10.4014/jmb.1904.04015.

[22] M. Myronovskyi, A. Luzhetskyy, Heterologous production of small molecules in the optimized Streptomyces hosts., Nat. Prod. Rep. 36 (2019) 1281–1294. https://doi.org/10.1039/c9np00023b.

[23] E. Angov, P.M. Legler, R.M. Mease, Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding., Methods Mol. Biol. 705 (2011) 1–13. https://doi.org/10.1007/978-1-61737-967-3_1.

[24] L. Pellizza, C. Smal, G. Rodrigo, M. Arán, Codon usage clusters correlation: towards protein solubility prediction in heterologous expression systems in E. coli., Sci. Rep. 8 (2018) 10618. https://doi.org/10.1038/s41598-018-29035-z.

[25] E. Angov, Codon usage: nature's roadmap to expression and folding of proteins., Biotechnol. J. 6 (2011) 650–659. https://doi.org/10.1002/biot.201000332.

[26] Z. Zhou, Y. Dang, M. Zhou, L. Li, C.-H. Yu, J. Fu, S. Chen, Y. Liu, Codon usage is an important determinant of gene expression levels largely through its effects on transcription., Proc Natl Acad Sci USA. 113 (2016) E6117–E6125. https://doi.org/10.1073/pnas.1606724113.

[27] N.A. Burgess-Brown, S. Sharma, F. Sobott, C. Loenarz, U. Oppermann, O. Gileadi, Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study., Protein Expr. Purif. 59 (2008) 94–102. https://doi.org/10.1016/j.pep.2008.01.008.

[28] Y. Wang, C. Li, M.R.I. Khan, Y. Wang, Y. Ruan, B. Zhao, B. Zhang, X. Ma, K. Zhang, X. Zhao, G. Ye, X. Guo, G. Feng, L. He, G. Ma, An engineered rare codon device for optimization of metabolic pathways., Sci. Rep. 6 (2016) 20608. https://doi.org/10.1038/srep20608.

[29] L.K. Edison, V.M. Dan, R. S. R, P. N. S, A Strategic Production Improvement of Streptomyces Beta Glucanase Enzymes with Aid of Codon Optimization and Heterologous Expression, Biosci., Biotechnol. Res. Asia. 17 (2020) 587–599. https://doi.org/10.13005/bbra/2862.

[30] H. Takahashi, T. Kumagai, K. Kitani, M. Mori, Y. Matoba, M. Sugiyama, Cloning and characterization of a *Streptomyces* single module type non-ribosomal peptide synthetase catalyzing a blue pigment synthesis., J. Biol. Chem. 282 (2007) 9073–9081. https://doi.org/10.1074/jbc.M611319200.

[31] K.J. Weissman, The structural biology of biosynthetic megaenzymes., Nat. Chem. Biol. 11 (2015) 660–670. https://doi.org/10.1038/nchembio.1883.

[32] C.R. Vickery, I.P. McCulloch, E.C. Sonnenschein, J. Beld, J.P. Noel, M.D. Burkart, Dissecting modular synthases through inhibition: A complementary chemical and genetic approach., Bioorg. Med. Chem. Lett. 30 (2020) 126820. https://doi.org/10.1016/j.bmcl.2019.126820.

[33] N.J. Hillson, R.D. Rosengarten, J.D. Keasling, j5 DNA assembly design automation software., ACS Synth. Biol. 1 (2012) 14–21. https://doi.org/10.1021/sb2000116.

[34] J.A. Myers, B.S. Curtis, W.R. Curtis, Improving accuracy of cell and chromophore concentration measurements using optical density., BMC Biophys. 6 (2013) 4. https://doi.org/10.1186/2046-1682-6-4.

[35] M. Wehrs, J.M. Gladden, Y. Liu, L. Platz, J.-P. Prahl, J. Moon, G. Papa, E. Sundstrom, G.M. Geiselman, D. Tanjore, J.D. Keasling, T.R. Pray, B.A. Simmons, A. Mukhopadhyay, Sustainable bioproduction of the blue pigment indigoidine: Expanding the range of heterologous products in *R. toruloides* to include non-ribosomal peptides, Green Chem. 21 (2019) 3394–3406. https://doi.org/10.1039/C9GC00920E.

[36] B. Pang, Y. Chen, F. Gan, C. Yan, L. Jin, J.W. Gin, C.J. Petzold, J.D. Keasling, Investigation of Indigoidine Synthetase Reveals a Conserved Active-Site Base Residue of Nonribosomal Peptide Synthetase Oxidases., J. Am. Chem. Soc. 142 (2020) 10931–10935. https://doi.org/10.1021/jacs.0c04328.

[37] M.-C. Lo, A. Aulabaugh, G. Jin, R. Cowling, J. Bard, M. Malamas, G. Ellestad, Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery., Anal. Biochem. 332 (2004) 153–159. https://doi.org/10.1016/j.ab.2004.04.031.

[38] A.M.J. Crowell, M.J. Wall, A.A. Doucette, Maximizing recovery of water-soluble proteins through acetone precipitation., Anal. Chim. Acta. 796 (2013) 48–54. https://doi.org/10.1016/j.aca.2013.08.005.

[39] Y. Chen, J. Gin, C. J Petzold, Discovery proteomic (DDA) LC-MS/MS data acquisition and analysis v2, (2021). https://doi.org/10.17504/protocols.io.buthnwj6.

[40] Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D.J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, S. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A.F. Jarnuczak, T. Ternent, A. Brazma, J.A. Vizcaíno, The PRIDE database and related tools and resources in 2019: improving support for quantification data., Nucleic Acids Res. 47 (2019) D442–D450. https://doi.org/10.1093/nar/gky1106.

[41] K. Huynh, C.L. Partch, Analysis of protein stability and ligand interactions by thermal shift assay., Curr. Protoc. Protein Sci. 79 (2015) 28.9.1-28.9.14. https://doi.org/10.1002/0471140864.ps2809s79.

[42] O. Pless, E. Kowenz-Leutz, G. Dittmar, A. Leutz, A differential proteome screening system for post-translational modification-dependent transcription factor interactions., Nat. Protoc. 6 (2011) 359–364. https://doi.org/10.1038/nprot.2011.303.

[43] P.C. Dorrestein, S.B. Bumpus, C.T. Calderone, S. Garneau-Tsodikova, Z.D. Aron, P.D. Straight, R. Kolter, C.T. Walsh, N.L. Kelleher, Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry., Biochemistry. 45 (2006) 12756–12766. https://doi.org/10.1021/bi061169d.

[44] S.C. Curran, A. Hagen, S. Poust, L.J.G. Chan, B.M. Garabedian, T. de Rond, M.-J. Baluyot, J.T. Vu, A.K. Lau, S. Yuzawa, C.J. Petzold, L. Katz, J.D. Keasling, Probing the

Flexibility of an Iterative Modular Polyketide Synthase with Non-Native Substrates *in Vitro*, ACS Chem. Biol. 13 (2018) 2261–2268. https://doi.org/10.1021/acschembio.8b00422.

[45] B. Schilling, M.J. Rardin, B.X. MacLean, A.M. Zawadzka, B.E. Frewen, M.P. Cusack, D.J. Sorensen, M.S. Bereman, E. Jing, C.C. Wu, E. Verdin, C.R. Kahn, M.J. Maccoss, B.W. Gibson, Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation., Mol. Cell. Proteomics. 11 (2012) 202–214. https://doi.org/10.1074/mcp.M112.017707.

[46] E.B. Vervoort, A. van Ravestein, N.N. van Peij, J.C. Heikoop, P.J. van Haastert, G.F. Verheijden, M.H. Linskens, Optimizing heterologous expression in dictyostelium: importance of 5' codon adaptation., Nucleic Acids Res. 28 (2000) 2069–2074. https://doi.org/10.1093/nar/28.10.2069.

[47] M.A. Collart, B. Weiss, Ribosome pausing, a dangerous necessity for co-translational events., Nucleic Acids Res. 48 (2020) 1043–1055. https://doi.org/10.1093/nar/gkz763.

[48] M.H. Moreira, G.C. Barros, R.D. Requião, S. Rossetto, T. Domitrovic, F.L. Palhano, From reporters to endogenous genes: the impact of the first five codons on translation efficiency in Escherichia coli., RNA Biol. 16 (2019) 1806–1816. https://doi.org/10.1080/15476286.2019.1661213.

[49] M.A. Skiba, F.P. Maloney, Q. Dan, A.E. Fraley, C.C. Aldrich, J.L. Smith, W.C. Brown, PKS-NRPS Enzymology and Structural Biology: Considerations in Protein Production., Meth. Enzymol. 604 (2018) 45–88. https://doi.org/10.1016/bs.mie.2018.01.035.

[50] M. Oresic, D. Shalloway, Specific correlations between relative synonymous codon usage and protein secondary structure., J. Mol. Biol. 281 (1998) 31–48. https://doi.org/10.1006/jmbi.1998.1921.

[51] C. Mignon, N. Mariano, G. Stadthagen, A. Lugari, P. Lagoutte, S. Donnat, S. Chenavas, C. Perot, R. Sodoyer, B. Werle, Codon harmonization - going beyond the speed limit for protein expression., FEBS Lett. 592 (2018) 1554–1564. https://doi.org/10.1002/1873-3468.13046.

[52] E. Angov, C.J. Hillier, R.L. Kincaid, J.A. Lyon, Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host., PLoS ONE. 3 (2008) e2189. https://doi.org/10.1371/journal.pone.0002189.