ACS SyntheticBiology

pubs.acs.org/synthbio Research Article

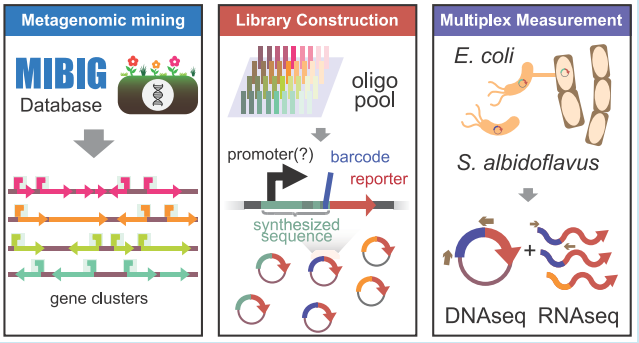
High-Throughput Transcriptional Characterization of Regulatory Sequences from Bacterial Biosynthetic Gene Clusters

Jimin Park, Sung Sun Yim, and Harris H. Wang*

Cite This: https://doi.org/10.1021/acssynbio.0c00639 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Recent efforts to sequence, survey, and functionally characterize the diverse biosynthetic capabilities of bacteria have identified numerous Biosynthetic Gene Clusters (BGCs). Genes found within BGCs are typically transcriptionally silent, suggesting their expression is tightly regulated. To better elucidate the underlying mechanisms and principles that govern BGC regulation on a DNA sequence level, we employed high-throughput DNA synthesis and multiplexed reporter assays to build and to characterize a library of BGC-derived regulatory sequences. Regulatory sequence transcription levels were measured in the Actinobacteria *Streptomyces albidoflavus* J1074, a popular model strain from a genus rich in BGC diversity. Transcriptional activities varied over 1000-fold in range and were used to identify key features associated with expression, including GC content, transcription start sites, and sequence motifs. Furthermore, we demonstrated that transcription levels could be modulated through coexpression of global regulatory proteins. Lastly, we developed and optimized a *S. albidoflavus* cell-free expression system for rapid characterization of regulatory sequences. This work helps to elucidate the regulatory landscape of BGCs and provides a diverse library of characterized regulatory sequences for rational engineering and activation of cryptic BGCs.



KEYWORDS: *synthetic biology, regulatory sequences, Streptomyces, biosynthetic gene clusters, cell-free expression systems, oligonucleotide library synthesis*

Beyond essential genes for core metabolism, many bacteria harbor genes to synthesize a variety of metabolites that can enhance fitness in different competitive environments.[1−4] These biosynthetic genes often localize in genomes in regions known as Biosynthetic Gene Clusters (BGCs), which encode proteins for regulating, synthesizing, and transporting secondary metabolites. Since these secondary metabolites also have many uses in medicine, agriculture, and industrial settings,[5,6] considerable efforts have been invested to discover novel BGCs and isolate their cognate biosynthetic products.[7−9] Genomic and metagenomic sequencing has dramatically increased the rate that BGCs are identified in microbes and across environments, which is further propelled by new computational tools and databases to accurately stratify BGC types and subtypes. While these advances have led to an exponential rise in the number of BGCs cataloged,[9−12] experimental characterizations and validations of BGCs have remained challenging. Under laboratory conditions, BGCs are generally transcriptionally silent in their native host or in a heterologous host, and the conditions to trigger their activation or tune their genetic regulation are poorly understood. Even if specific molecules are produced from BGCs, they often can have cryptic or unknown functions that may require further detailed and labor-intensive studies.[13]

Numerous strategies have been developed to activate "silent" BGCs.[13] Testing environmental stimulatory signals or triggers through growth condition optimization has been a useful and sometimes fruitful approach. For instance, altering growth media compositions, treating populations with chemical perturbations, or coculturing BGC-harboring microbes with other microbes have all demonstrated some level of success to induce BGC expression and elicit secondary metabolite synthesis.[14−17] Another strategy is to take BGCs from challenging native hosts and transform, express, and screen them in a heterologous host more amenable to laboratory cultivation and analysis. This heterologous expression approach enables high-throughput screening of diverse and complex metagenomic DNA samples (*e.g.*, environmental DNA) for their biosynthetic potential.[18−21] Synthetic biology also offers a pathway refactoring approach where synthetic

Received: December 22, 2020

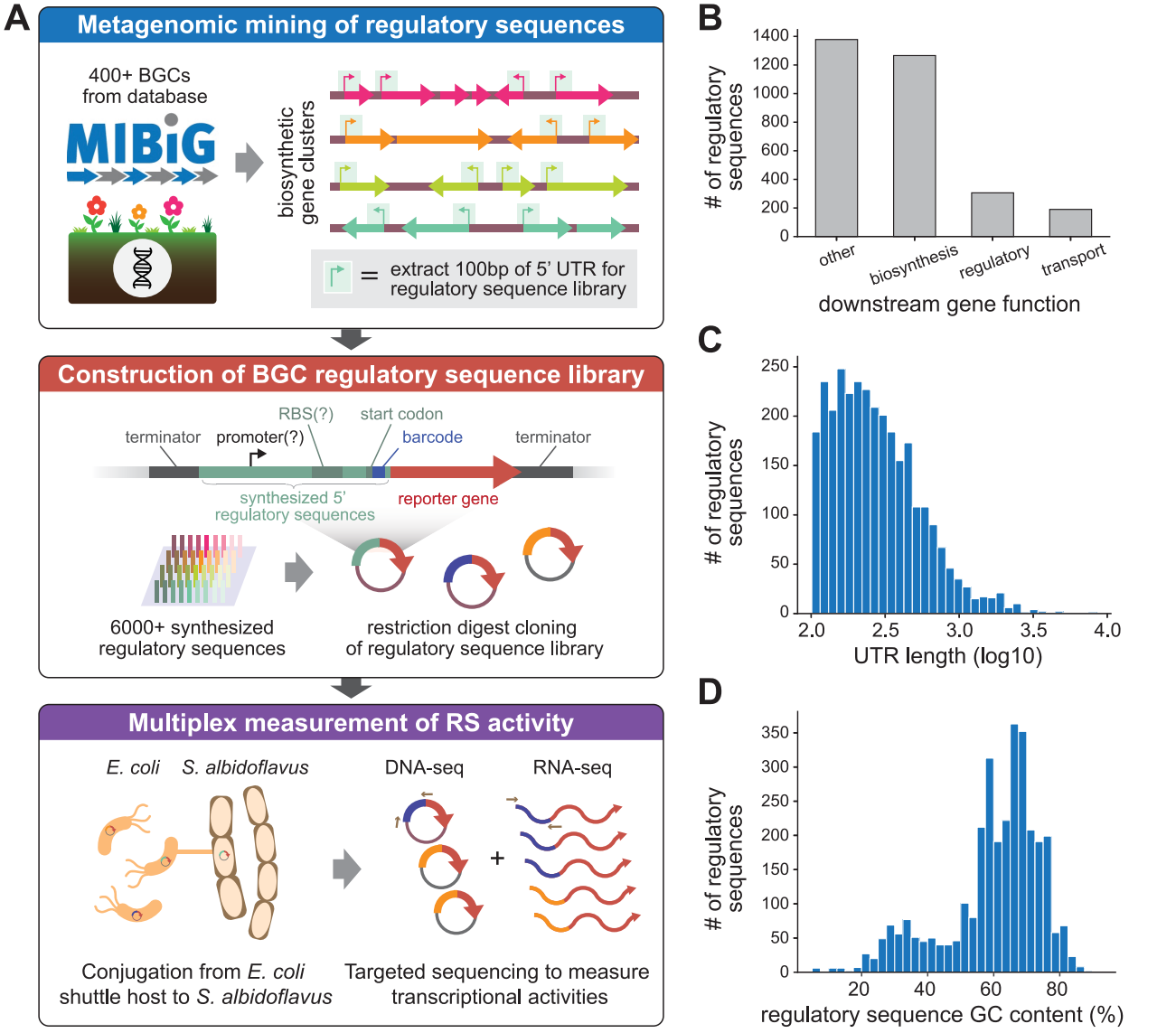


Figure 1. Mining and construction of a regulatory sequence library from Biosynthetic Gene Clusters. (A) Overview of our pipeline to characterize transcriptional activities of regulatory sequences in *S. albidoflavus* J1074. Regulatory sequences are mined from intergenic regions of MIBiG BGCs. Subsequently, a start codon, unique barcode sequence, and restriction digest sites are added to each sequence and all compiled sequences are synthesized together as an oligonucleotide pool. Following cloning in *E. coli* and subsequent conjugation into *S. albidoflavus*, targeted DNA-seq and RNA-seq are used to measure transcriptional activities of each library sequence. (B) Different gene function categories associated with CDSs downstream of mined intergenic regions. Categories were assigned by their smCOG designations. (C) Distribution of length of intergenic regions where regulatory sequences were mined in BGCs. (D) Distribution of the GC content within the first 100 base pair of the regulatory sequences mined from BGC intergenic regions.

versions of natural BGCs are synthesized *de novo* with well-characterized promoters and ribosomal binding site (RBSs) sequences replacing the native regulatory elements and all genes are codon optimized to improve heterologous expression. While these efforts are resource-intensive, they have shown recent successes in generating new secondary metabolites.[22−25]

Due to the notable metabolic burden of producing secondary metabolites, BGCs are generally tightly regulated and their gene expression control offers an interesting target for direct BGC activation. For instance, rewiring gene regulatory networks by tuning the expression of key global regulatory genes could in turn activate BGCs.[26] Modulating expression of regulatory genes, including both pathway/cluster specific regulators and broad global regulators, have been shown to induce expression of cryptic BGCs and synthesis of novel secondary metabolites.[27−30] These studies suggest that rewiring regulatory networks may potentially be a simple and generalizable approach to activate secondary metabolism. A deeper profiling of BGC gene regulation could therefore accelerate the secondary metabolite discovery pipeline.

Massively Parallel Reporter Assays (MPRA) have emerged as a powerful tool to systematically dissect gene regulation.[26,31−33] By leveraging high-throughput DNA oligonucleotide library synthesis, MPRAs enable systematic and functional characterization of a large and diverse sequence space for elucidating fundamental principles that drive gene regulation and expression. Here, we describe the use of this general approach to assess gene expression determinants of BGC regulatory sequences. We employed MPRA to characterize a

https://doi.org/10.1021/acssynbio.0c00639
ACS Synth. Biol. XXXX, XXX, XXX−XXX

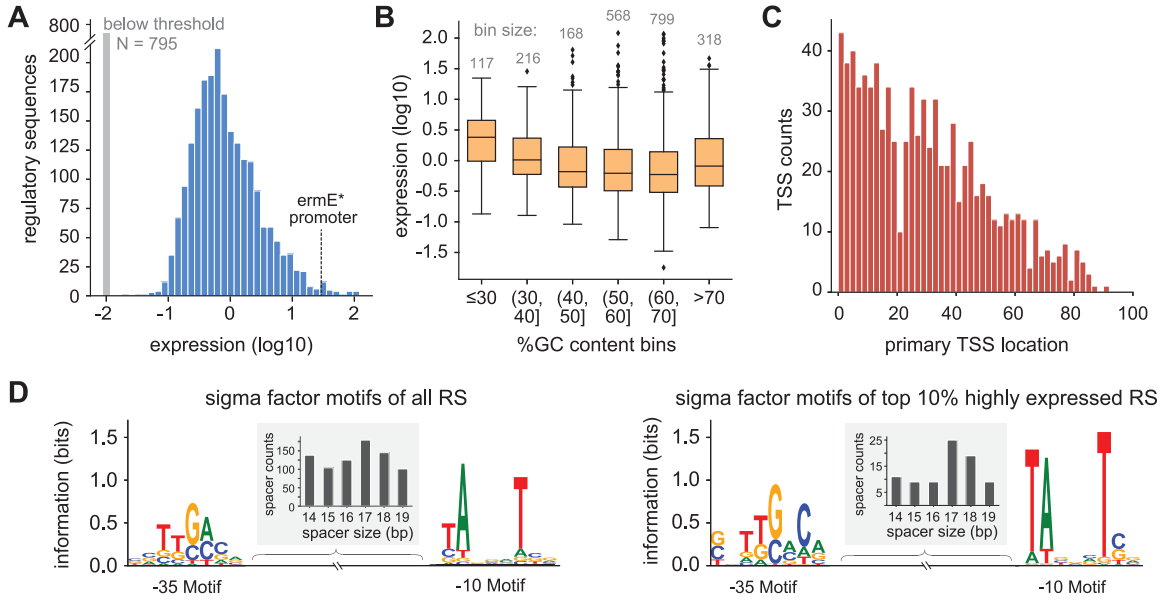


Figure 2. Transcriptional activity of BGC-derived regulatory sequences (RS) in *S. albidoflavus*. (A) Histogram showing the distribution of transcriptional activity of 2186 regulatory sequences in *S. albidoflavus* is displayed in blue. 795 RSs below the expression detection threshold (*i.e.,* observed >10 DNA reads but >10 RNA reads) were given a pseudovalue and displayed in gray. Expression level of the constitutive *Streptomyces* promoter ermE* is denoted by the dotted line. (B) Box plot showing transcriptional activity of RS grouped by their GC content. Center line denote the median, with boxes showing top and bottom 25th percentiles. Whiskers are 1.5× the respective boxes and values beyond the whiskers are displayed as individual points. (C) Distribution of primary TSS locations of 853 regulatory sequences measured in *S. albidoflavus*. (D) Sigma factor motifs generated from regulatory sequences upstream of TSS from all active regulatory sequences (left panel) and the top 10% highest expression regulatory sequences (right panel). Between each −35 and −10 sequence motif, the distributions of spacer lengths are shown.

library of >3000 natural BGC regulatory sequences using *Streptomyces albidoflavus* J1074 (previously *Streptomyces albus* J1074[34]) as a model host. This regulatory library was transcriptionally active and exhibited expression ranges spanning >1000-fold that was correlated with GC content in the regulatory sequences (RSs). In addition, coexpression of global regulatory factors altered the expression capacity of different regulatory sequences. To accelerate MPRA studies in *Streptomyces*, we further developed and optimized a *Streptomyces* cell-free expression system capable of single-pot multiplex transcriptional characterization of regulatory sequences. These efforts offer a more systematic analysis of BGC regulation and a large set of characterized regulatory sequences that can be leveraged to better modulate and activate the expression of BGCs and other synthetic biology efforts in Actinobacteria.

RESULTS AND DISCUSSIONS

Mining and Generating a BGC Regulatory Sequence Library in a Model Actinomycetes. Since most regulatory sequences (RSs) in BGCs remain poorly characterized, we sought to perform an unbiased systematic survey of transcriptional capacities of different regulatory sequences found in BGCs. We chose *S. albidoflavus* J1074 as the bacterial host to characterize regulatory sequences in because this strain has been a useful host for heterologous expression of BGCs and is a model Actinobacteria from a genus rich in BGC diversity.[35,36] In order to facilitate screening a large set of diverse regulatory sequences, we leveraged MPRAs, which have been used extensively in bacteria for diverse applications ranging from characterizing diverse metagenomic regulatory sequences, assaying regulatory sequence composability, dissecting key sequence features underlying gene expression, and more.[31−33] Furthermore, while regulatory sequence libraries have been screened in *Streptomyces* before,[37−39] the transcriptional landscape of native BGC regulatory sequences have not been explored. To that end, we first mined *in silico* putative regulatory sequences from untranslated intergenic regions upstream of coding genes in BGCs, which were then synthesized in a pooled library, cloned into expression vectors, and then introduced into *S. albidoflavus* for MPRA measurements (Figure 1A). Our library of regulatory sequences was compiled from ∼400 full length BGCs in the Minimum Information about a Biosynthetic Gene cluster (MIBiG) database.[40] These curated BGCs produced a diverse array of secondary metabolite classes and were sourced from microbes with diverse phylogenetic origins (Figure S1A). Since Actinobacteria are prolific secondary metabolite producers,[1] the largest proportion of BGCs were derived from Actinobacteria genomes. Polyketides constituted the largest fraction of the curated BGC diversity followed by nonribosomal peptides (NRPs) and ribosomally synthesized and post-translationally modified peptides (RiPPs), respectively. The compiled BGCs ranged from 1 kilobase (kb) to 150 kb in size, with an average length of ∼41 kb (Figure S1B) and encoding ∼20 annotated coding sequences (CDS) (Figure S1C) with about half of the CDSs carrying gene function annotations. While most of the CDSs were biosynthesis genes, a minor subset of CDSs were genes related to regulatory and transport functions (Figure 1B). From these BGCs, we extracted 3189 5′ intergenic regions that were at least 100 base pairs (see Materials and Methods). These intergenic regions had an average length of ∼360 bp and majority of them were shorter than 1000 bps (Figure 1C). On average, each BGC yielded an average of eight valid 5′ intergenic regions. From each 5′ intergenic region, we selected 100 base pairs from the 3′ end, which corresponded to regions directly upstream of start codons used

https://doi.org/10.1021/acssynbio.0c00639
ACS Synth. Biol. XXXX, XXX, XXX−XXX

to identify 5′ intergenic regions, to yield a 3189-member library composed of 100 bp putative regulatory sequences (sequences that encoded a *Bam*HI or *Pst*I restriction cut sites in the 100 bp region were removed). Since regulatory sequence GC content is known to correlate with bacterial gene expression,[31,41] we also assessed the GC content of our regulatory sequences and observed two discrete clusters of GC content distribution: a major peak centered around ∼65% GC and a minor peak centered around ∼35% GC reflected the genomic GC content bias of known bacteria that are prolific secondary metabolite producers[40,42] (Figure 1D).

For each member of the regulatory sequence library, we assigned two unique 12-mer DNA barcode tags to yield a final library of 6378 constructs for pooled oligonucleotide library synthesis and multiplex reporter assays. The final library also included an inducible ptipA,[43] constitutive ermE*,[44] and a panel of endogenous *S. albidoflavus* promoters previously identified to be constitutively expressed[38] as control regulatory sequences. Each construct was synthesized with an ATG start codon, aforementioned DNA barcode, *Bam*HI and *Pst*I restriction digestion sites for cloning, and common flanking sequences for pooled PCR amplification. The synthesized oligonucleotide library pool was amplified and restriction digest cloned upstream of an ATG-less mCherry reporter gene in the pJP50 shuttle vector (a pIJ10257[45] derivative, see Materials and Methods) with the ΦBT1 integrase. We transformed the cloned library first into *E. coli* S17 strain and then used bacterial conjugation to deliver the library from *E. coli* S17 to *S. albidoflavus* J1074. Sequencing of the regulatory library in *S. albidoflavus* validated that ∼6000 unique RS-barcode constructs were successfully integrated into the *S. albidoflavus* genome, representing a ∼90% coverage of the designed library.

Multiplexed Measurements of BGC Regulatory Sequences in *S. albidoflavus*. To quantify the transcriptional capacity of our BGC regulatory sequence library in *S. albidoflavus*, we performed targeted DNA-seq and RNA-seq on the cell population upon growth in rich media. After collapsing down on the replicate DNA barcodes, 2981 putative regulatory sequences were successfully integrated into the *S. albidoflavus* genome (Figure S2A). Of those, 2186 sequences exhibited measurable transcriptional activity levels while the other 795 sequences had transcriptional levels below detection threshold (detectable DNA reads but no RNA reads; see Materials and Methods). Briefly, transcriptional activity levels were determined by normalizing the relative RNA abundance by the relative DNA abundance for each regulatory sequence. Individual transcriptional measurements were highly reproducible between biological replicates, and to a lesser extent, between barcode replicates (same regulatory sequences measured with two unique DNA barcode sequences) (Figure S2B). Transcriptional levels spanned over a 1000-fold range of expression, indicating that the BGC regulatory sequences contained a transcriptionally diverse set of regulatory sequences (Figure 2A). Comparing the transcriptional levels of BGC regulatory sequences against the widely used constitutive *Streptomyces* promoter, ermE*, showed that the vast majority of BGC regulatory sequences had lower transcriptional levels than ermE* (Figure 2A). Only 24 regulatory sequences (approximately 1% of measured regulatory sequences) exhibited higher transcriptional level than ermE*. In addition, none of the 24 sequences outperformed ermE* expression by more than 5-fold (most were between 2−4-fold higher than ermE*), indicating that expression levels of regulatory sequences derived from BGCs are generally weaker in *S. albidoflavus* than a common constitutive promoter ermE*. This observation suggests that BGC expression could be improved by using heterologous regulatory sequences such as ermE* or other characterized regulatory sequences with higher expression than ermE*.[37,38]

Next, to better contextualize the distribution of BGC regulatory sequence expressions, we explored the relationship between transcription levels and various global properties, such as GC content, phylogenetic origins and types of BGCs from which the regulatory sequences were derived. Regulatory sequence GC content has been noted to inversely correlate to gene expression in many other bacteria, with low GC content sequences exhibiting higher expression levels.[41] In agreement with previous findings, we also observed that BGC regulatory sequence expression was inversely correlated to GC content (Figure 2B). Accordingly, we observed that regulatory sequences with measured transcriptional levels, in contrast with sequences with transcriptional levels below the level of detection (*i.e.*, inactive or silent regulatory sequences), had lower average GC content (Figure S2C). In addition to GC content, we also compared transcriptional expression level distributions of regulatory sequences partitioned by either their phylogenetic origins, chemical compound type of their respective BGC metabolite, or functional class of their downstream gene. While a few transcriptional measurement distributions exhibited statistically different averages, none of the averages differed by more than an order of magnitude, suggesting that no categorical classifications could organize regulatory sequence gene expression by their transcriptional levels (Figure S2D). Therefore, to better identify gene expression determinants, we next dissected the relationship between expression and DNA sequence features at the base pair level.

Core bacterial transcription is mediated by sigma factors that interact with a bipartite cis-regulatory DNA sequence motif, commonly known as the −10 and the −35 sequence motifs. In order to map sigma factor binding locations in the BGC regulatory sequence library, we first compiled transcription start sites (TSS) of regulatory sequences using a k-means clustering approach as previously described.[31,46] TSS locations were highly correlated between biological replicates ($r > 0.98$) and barcode replicates ($r > 0.95$) (Figure S3A,B) and therefore all replicate data were pooled and the TSSs were compiled to yield 853 primary TSS locations (k-means clusters containing ≥70% of TSS calls). In contrast to the ∼2100 transcriptional level measurements, a fewer number of primary TSS locations were identified due to the relatively higher number of RNA reads required for robust TSS mapping (see Materials and Methods). The bulk of primary TSS locations were found within 50 bp from the start codon (Figure 2C), in agreement with TSS locations of native regulatory sequences in *S. coelicolor*.[47] Interestingly, we also noted the high prevalence of leaderless transcripts in *Streptomyces* species (∼24% of primary TSS were mapped to ≥10 bp).[47] A subset of regulatory sequences with primary TSS (∼41%) appeared to have multiple TSS locations and this observation correlated with GC content; regulatory sequences with more than one TSS locations had significantly lower regulatory sequence GC content than sequences with a single TSS location (*t* test, *p*-value <0.005) (Figure S3C,D).

https://doi.org/10.1021/acssynbio.0c00639
ACS Synth. Biol. XXXX, XXX, XXX−XXX

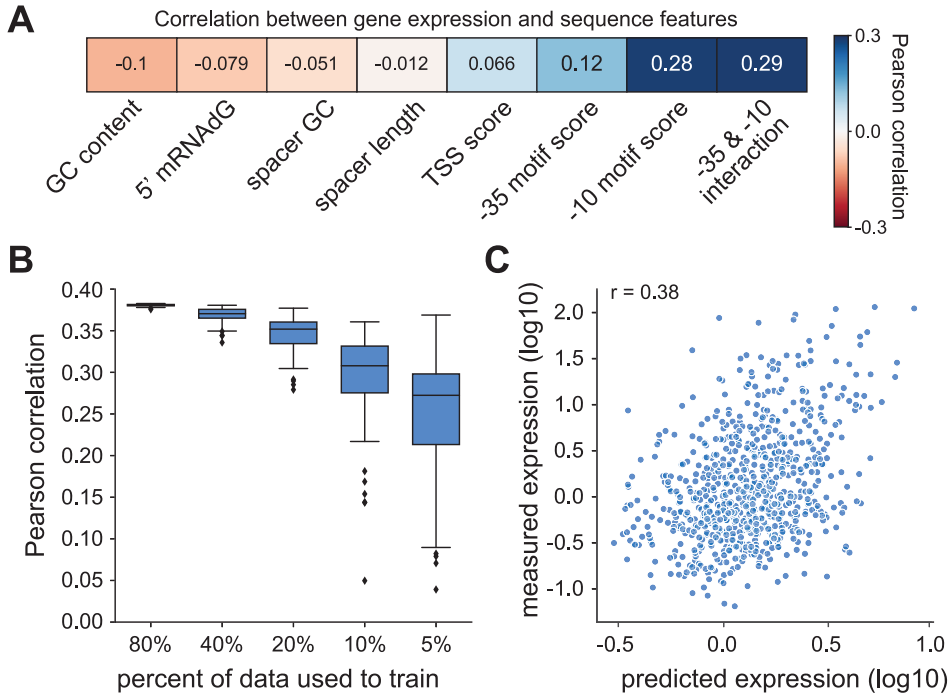


Figure 3. Modeling the transcriptional activity of BGC-derived regulatory sequences using a multiple linear regression. (A) Pearson correlation between measured gene expression and various sequence features for 835 BGC regulatory sequences. (B) Correlation between predicted and measured gene expression levels for varying percentages of the data used for training and testing sets. (C) Example multiple linear regression model using 80% of the data set as training and 20% of the data as the test set.

Next, using primary TSS locations for each regulatory sequence, we analyzed the upstream regulatory regions to identify sigma factor binding sites. Using BioProspector,[48] we identified an enriched bipartite sequence motif (ttgacn/TAnnnT) (Figure 2D) similar to the canonical sigma factor motif from *E. coli* (TTGACA/TATAAT). Interestingly, between *E. coli* and *S. albidoflavus*, both of their −10 and the −35 sequence motifs are similar, but their −10 motifs appear to be stronger than their −35 motifs. In the −35 motif from *S. albidoflavus*, −33 G has the highest signal, similar to the primary sigma factor, HrdB, binding motifs identified from *Streptomyces coelicolor* RNA-seq results (ntGacn/tAnnnT).[47,49] In addition, the most common spacing between the −10 and −35 motifs was 17 bp, which reflects the optimal canonical sigma factor motif spacings.[50] Next, to identify sequence features associated with high gene expression, we selected a subset of sequences with top 10% highest transcriptional levels and searched for sequence motif, which yielded a similar but slightly different motif (Figure 2D). For highly expressed sequences, the entire −35 motif was generally more conserved. Specifically, the TTG motif was more prominently in addition to the −31 C base. For the −10 motif, on the other hand, the −12 T base was stronger in highly expressed regulatory sequences. Lastly, within this subset of high expression sequences, we observed that a higher proportion of sequences had the 17 bp canonical optimal spacing between the −10 and −35 motifs (Figure 2D).

Given that many different sequence properties (*i.e.*, −10/−35 sequence motif and interaction, GC content, spacer sequence, 5′ mRNA stability, and TSS region composition)[46] can affect gene expression, we attempted to integrate these features into a model of gene expression in *Streptomyces albidoflavus* using a subset of 835 RSs. First, each sequence feature was independently compared against transcription level. Similar to previous reports,[31,46,51] we find in general that the −10 motif strength and the −10/−35 sequence motif interaction score showed the strongest correlation to gene expression (Figure 3A). Next, these sequence features were combined to generate a multiple linear regression (MLR) model to capture the relationship between sequence features and transcriptional activity levels. This model was subjected to multiple rounds of subsampling, training, and cross-validation, and we found that the linear model trained on 80% of the data could explain ∼14% of the observed variance (r = 0.38) on average in the other 20% of the data (Figure 3B,C). We note that similar previous MLR models, which were trained on expression data from similar experiments but different regulatory sequence library and in bacterial species, could explain a higher proportion of variance,[46] which suggests that BGC RS expression in this particular microbe may rely on other regulatory expression mechanisms or utilize sequence features not considered in our MLR model.

Modulation of BGC Regulatory Sequence Expression Using Global Regulatory Proteins. Having characterized key sequence properties that underlie expression of BGC cis-regulatory sequences, we next aimed to dissect the trans-regulatory mechanisms that drive BGC expression. Global regulatory proteins have pleiotropic roles in bacterial gene regulatory networks, capable of drastically rewiring the cellular transcriptome and modulating gene expression at many loci. As such, global regulatory proteins could be potent factors that, when coexpressed, may activate secondary metabolism and improve natural product yields.[9] We thus sought to determine whether heterologous expression of global regulatory factors could induce direct transcriptional changes in our BGC-derived regulatory sequence library. To that end, we first assembled a list of candidate global regulatory proteins that were previously used to improve natural product yields in

https://doi.org/10.1021/acssynbio.0c00639
ACS Synth. Biol. XXXX, XXX, XXX−XXX

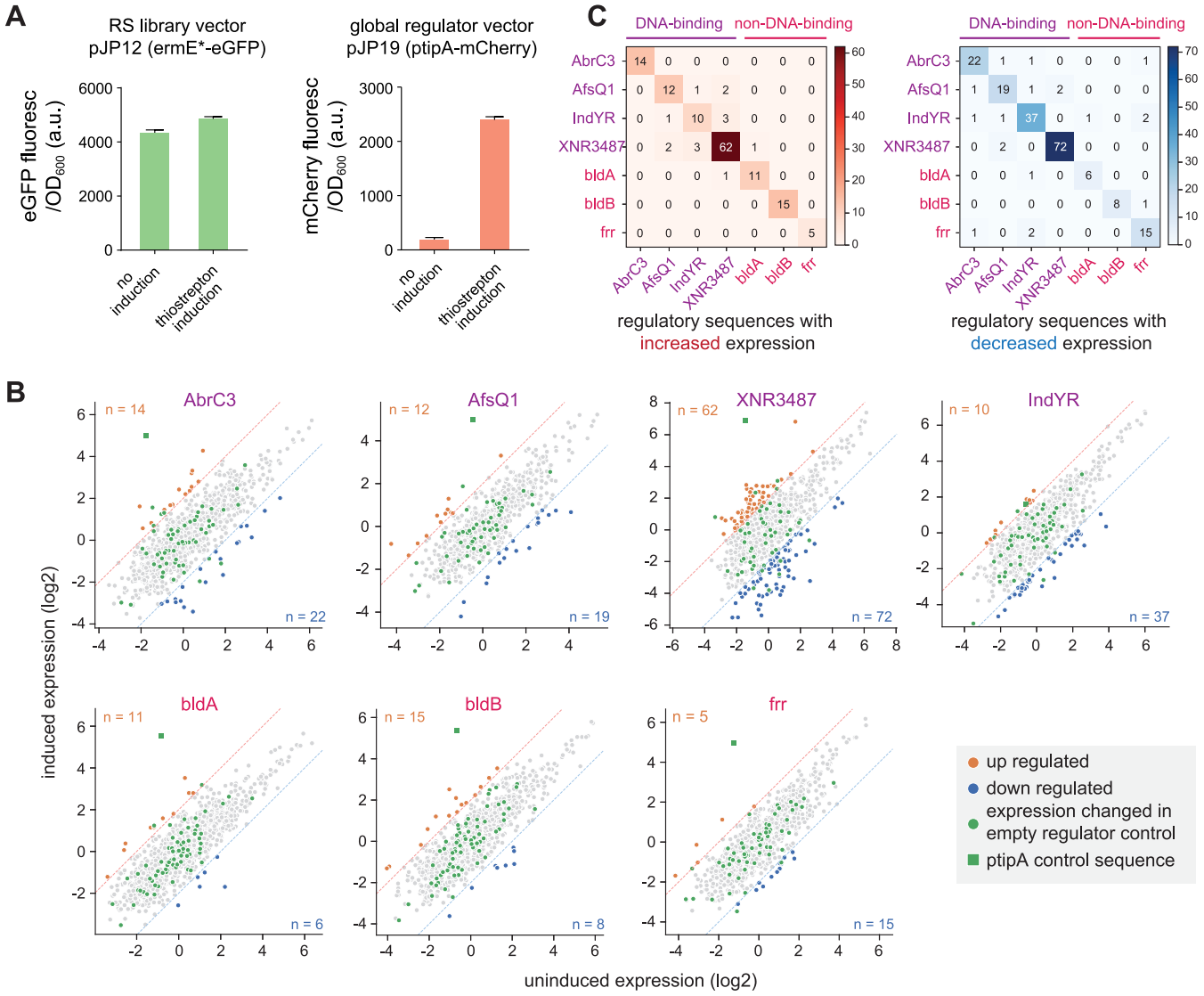


Figure 4. Modulating transcriptional activities of regulatory sequences through expression of global regulatory proteins. (A) Normalized activity of two fluorescent reporter genes in *S. albidoflavus* showing comaintenance and fluorescent reporter expression from two separate plasmid constructs using two different integration systems (pJP12, which shares vector backbone for the RS library and integrates through ΦBT1; pJP19, which shares vector backbone for the global regulator constructs and integrates through ΦC31) measured through a microplate reader. (B) Transcriptional activity between induced and uninduced samples across each regulator-RS library. Orange and blue data points represent RS with more than 4-fold differential expression compared to uninduced condition. The positive control of induction, ptipA, is displayed in green square. Sequences that had altered expression changes are represented in green circles. (C) Heatmaps showing the number of RS with altered transcriptional activity (increase on left panel and decrease on right panel) for each global regulator as well as RS that can be modulated by more than one regulator.

different *Streptomyces* species (Table S1). DNA-binding global regulators we explored included IndYR, AbrC3, AfsQ1, and XNR3487. lndYR, a GntR-like transcription factor, is an endogenous factor in *S. globisporus* that regulates the expression of landomycin E gene cluster and also yields higher titers of actinorhodin (ACT) and moenomycin when heterologously expressed in *S. coelicolor* and *S. ghanaensis*, respectively.[52−54] The two-component system (TCS) response regulator (RR) AbrC3 could activate the natively silent ACT gene cluster when heterologously expressed in *S. lividans*.[55,56] Furthermore, AbrC3 also increased synthesis of desferrioxamine when overexpressed in *S. agrillaceus*.[57] AfsQ1, another TCS RR, exhibits pleiotropic phenotypes when heterologously expressed; in *S. lividans*, AfsQ1 increased ACT and undecylprodigiosin antibiotic production, while in *S. coelicolor*, it improved coelimycin production and altered phosphate metabolism and morphological development.[58−61] The endogenous TCS RR from *S. albidoflavus*, XNR3487, did not have any experimentally verified function but has high sequence homology to *S. coelicolor* CseB (87% sequence identity), which is a regulator of extracellular stress response pathway.[62] CseB is known to regulate the expression of sigE, a sigma factor that has been shown to modulate secondary metabolism expression in response to cell wall damage.[63−65] In addition, we also explored non-DNA binding global regulators documented to induce BGC activation, including *bldA*, *bldB*, and *ffr*. Notably, *bldA*, a tRNA gene that translates a rare *Streptomyces* codon, TTA, is only expressed in the late morphological development phase and genes required for secondary metabolism frequently encode the TTA rare codon.[66,67] BldB is a small 98 AA protein that is indispensable for aerial mycelia formation and antibiotic synthesis in

https://doi.org/10.1021/acssynbio.0c00639
ACS Synth. Biol. XXXX, XXX, XXX−XXX

S. coelicolor with little known about its mechanism of action.[68−70] Lastly, *frr* is a gene encoding the Ribosome Recycling Factor that increased toyomycin synthesis when overexpressed in *S. diastatochromogenes*.[71−73] Other global regulatory factors were considered (AdpA, AtrA, crp, metK, and NdgR), but our attempts to generate the necessary expression constructs were unsuccessful. Together, we explored seven diverse global regulators for their capacity to induce expression changes in our BGC-derived regulatory sequence library.

To test and screen each regulator for their impact on our RS library, we first generated different *S. albidoflavus* strains that could be induced to overexpress each global regulator, then transformed the RS library into each strain, and finally measured the transcriptional levels of the RS library by deep sequencing as before. Briefly, we cloned each regulator downstream of a thiostrepton inducible promoter in a pIJ6902 shuttle vector and conjugated the plasmids from *E. coli* to *S. albidoflavus* where they integrated into the genomes using the ΦC31 integrase system. We then performed library conjugations, *via* the ΦBT1 integration system, into each regulator strain to yield eight different *S. albidoflavus* library-regulator populations (seven regulator strains and one strain with an empty pIJ6902 vector as a control). The two integration systems, ΦBT1 and ΦC31, are reported to be compatible in other *Streptomyces* strains.[74] We confirmed compatibility in *S. albidoflavus* through sequential integration of two plasmid constructs, pJP12 which encodes a constitutive ermE*-eGFP and integrates *via* ΦBT1, and pJP19 which encodes an inducible ptipA-mCherry and integrates *via* ΦC31, and subsequent fluorescence measurements (Figure 4A). Regulatory sequence abundance distributions of each library populations were similar to that of the original RS library constructed in the wild-type *S. albidoflavus* (Figure S4A). To profile transcriptional activation of regulatory sequences with and without global regulator expression in parallel, library populations were cultured with and without thiostrepton induction and targeted DNA and RNA-seq was performed as before. Transcriptional levels from these regulator-RS library samples had similar correlations between biological, technical, and barcode replicates as the original RS library samples (Figure S4B). However, we noted that fewer regulatory sequences had detectable transcription; the initial RS library had ∼2200 measurable sequences while regulator-RS library samples had ∼1000 measurable sequences on average (Figure S4C). Rarefaction curves of RNA samples from the regulator-RS library suggested that differences in sequencing coverage likely does not explain the observed discrepancies between the number of measurable sequences (Figure S4D). The differences in the number of measurable sequences between the libraries may reflect the metabolic burdens elicited by additional antibiotic selection required for the regulator expression construct (further decrease in measurable sequences were also observed in thiostrepton induced regulator-RS library samples). Overall, these results highlight the feasibility of assessing RS library activity in the context of different global regulatory proteins *via* pooled MPR assays.

To enable robust transcriptional comparisons between samples, we median normalized each sample and compared expression of each regulator-RS library in induced and uninduced conditions (Figure S4E). First, we confirmed thiostrepton induction in all library populations by assessing activation of a control thiostrepton-inducible ptipA promoter present in each library. Indeed, we detected induction of ptipA promoter across all samples (ranging from 4-fold to 300-fold increase) (Figure S4F). 4-fold induction was used as a threshold to designate RS as those with altered expression levels (up or down regulated) from nonresponsive regulatory sequences. We further filtered out any regulatory sequences that also showed differential expression in the empty vector library (*i.e.*, "no regulator") to control for any effects by the vector itself (Figure S4G). On average, heterologous expression of a global regulator resulted in about 18 RS upregulated and 22 RS downregulated on average, which is ∼1% of the RS library (Figure 4B). Overall, there were 279 unique sequences with altered expression spanning 169 distinct BGCs. The ptipA sequence was the most differentially expressed sequence for each library, averaging a ∼100-fold increase; upregulated RS had an average of ∼5.8-fold increase, while downregulated RS had ∼5.6-fold decrease in expression, reflecting the relatively weak magnitudes of gene expression changes. Interestingly, the endogenous *S. albidoflavus* TCS RR XNR3487 had the largest number of differentially altered RS (65 downregulated and 77 upregulated), suggesting that regulatory network rewiring with endogenous factors may induce the largest gene expression changes. Furthermore, majority of gene expression changes were orthogonal with minimal overlap of altered sequences between regulators, which suggests that expression of different global regulators could differentially modulate different BGC regulatory sequences (Figure 4C). Lastly, we observed that DNA binding regulatory factors generally yielded more differentially expressed sequences (average of ∼38 downregulated and ∼25 upregulated) compared to non-DNA binding pleiotropic global regulators (average of ∼10 upregulated and ∼9 downregulated), highlighting that as one may expect, the non-DNA binding global regulators are less likely to activate secondary metabolism through transcriptional regulation than DNA-binding factors.

An *in Vitro S. albidoflavus* Transcription System for Regulatory Activity Measurements. A key bottleneck for multiplexed, library scale transcriptional assays such as the ones demonstrated in this study is the need for high-efficiency transformation protocols. Recently, cell-free expression systems have emerged as a powerful platform for rapid and high-throughput characterization of regulatory sequences.[75] Cell-free extracts (*i.e.*, cell lysates) prepared from cultured cells contain native transcriptional machinery and can recapitulate cellular transcriptional profiles. As such, cell-free expression systems are especially suited for profiling regulatory sequence libraries in nonmodel microbes where efficient transformation protocols are not available or are difficult to carry out at the necessary scale. Previously, several reports described cell-free *Streptomyces* lysates that use T7 promoters and purified T7 RNA polymerases to drive transcription.[76−78] Encouraged by these successes, we sought to develop a novel cell-free expression system for *S. albidoflavus* for multiplexed transcriptional reporter assays with the aim of establishing a generalizable platform reliant on native transcriptional machinery for rapid and massively parallel transcriptional measurements that are concordant with *in vivo* transcriptional measurements. To that end, we used a previous approach we described for generating simple bacterial cell lysate systems.[46] Briefly, cells were cultivation in a liquid medium, lysed through sonication, incubated in a runoff reaction, and dialyzed (Figure

https://doi.org/10.1021/acssynbio.0c00639
ACS Synth. Biol. XXXX, XXX, XXX−XXX

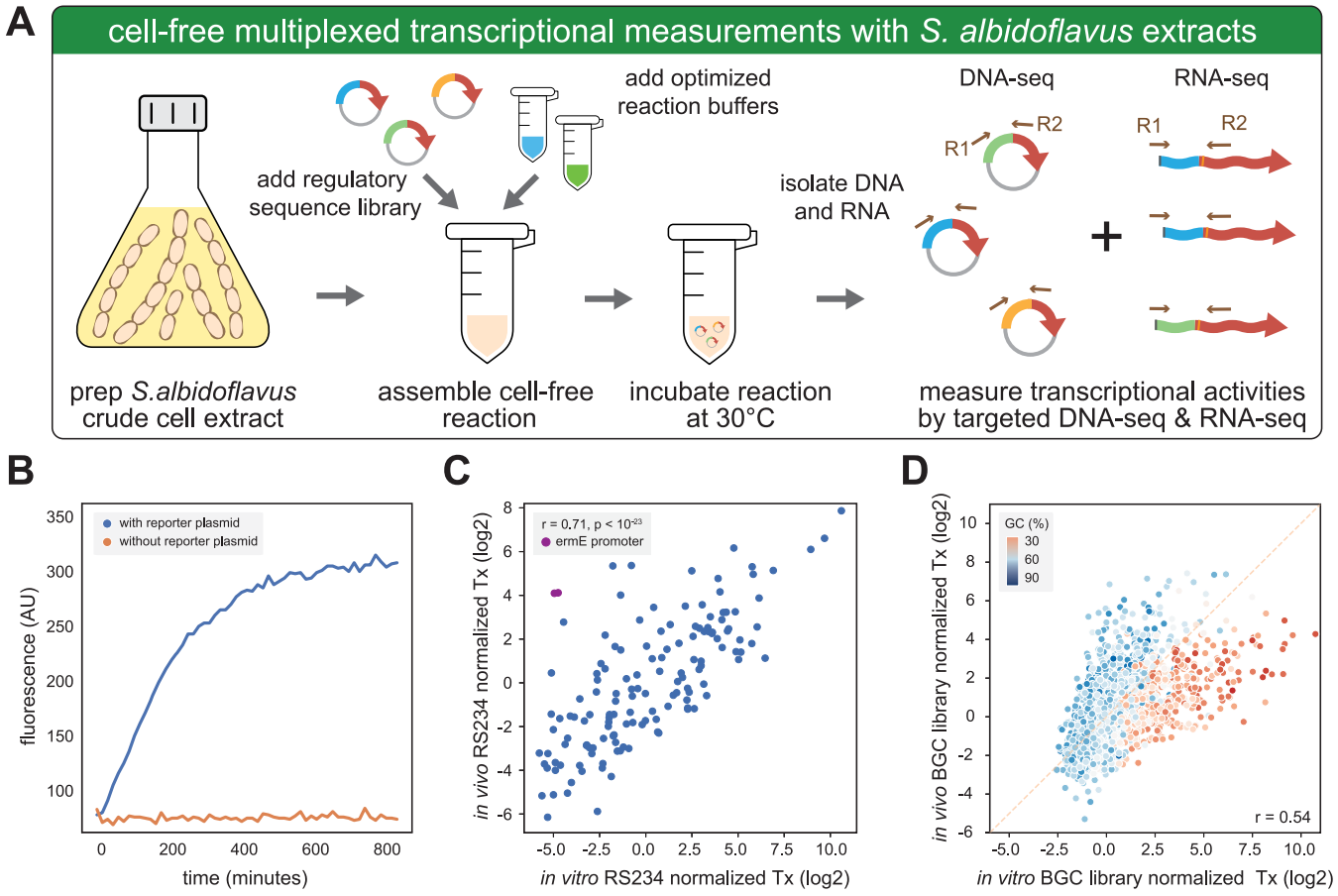


Figure 5. Cell free expression platform for rapid characterization of regulatory sequence activities in *S. albidoflavus*. (A) Overview of generating a *S. albidoflavus* cell-free expression system to measure transcriptional activities of regulatory sequence libraries. Following preparation of cell extracts, cell-free reaction is assembled by mixing cell lysates with optimized buffers and plasmids containing the RS libraries. The reaction is then incubated, and then separate DNA and RNA sequencing libraries are prepped and sequenced. (B) Transcriptional assay with *S. albidoflavus* cell-free lysate using a Broccoli reporter construct (blue) and a negative control (orange). (C) Quantification of *S. albidoflavus* cell-free expression system with an established set of 234 regulatory sequences (RS234). The constitutive promoters ermE* with two separate barcodes are denoted by purple circles. (D) Correlation between *in vivo* and *in vitro* transcription measurements on the BGC-derived RS library. GC content of the regulatory sequences are denoted in different colors.

5A). Buffer concentrations of Mg-glutamate and K-glutamate, which are important cell-free reaction substrates,[79] were optimized using Broccoli, an RNA aptamer that fluoresces when bound to DHFBI-1T, as a reporter. We found that our optimized *S. albidoflavus* cell-free lysate could robustly producing a Broccoli signal compared to a control construct (Figure 5B, Figure S5A).

Using a previously described set of 234 regulatory sequences (RS234) quantified in different cell-free expression systems,[46] we then measured transcriptional activities both *in vivo* and *in vitro* in *S. albidoflavus* and found that the two measurements were reasonably well correlated (r = 0.71) (Figure 5C). Interestingly, the strong constitutive promoter ermE* *in vivo* exhibited low transcriptional activity *in vitro*, in accordance with our observation that the *S. albidoflavus* extract did not yield measurable Broccoli signal when the reporter was driven by an ermE* promoter (data not shown). Nevertheless, we next measured the transcriptional activity of our BGC RS library using the *S. albidoflavus* lysate and compared the *in vitro* activity to *in vivo* data. Overall, we observed a weaker but notable correlation (r = 0.54) between *in vitro* and *in vivo* transcriptional measurements (Figure 5D). We found that GC content of the RS more strongly influenced gene expression *in vitro* (r = −0.59) than *in vivo* (r = −0.1) (Figure S5B). RS with low GC content had higher transcription *in vitro* than *in vivo*, while high GC RS had higher transcription *in vivo* than *in vitro* (Figure 5D). These results suggest that differences in transcriptional capacity in *S. albidoflavus* cellular and cell-free environments may yield notable differences in gene expression measurements while still conserving general expression patterns, and that further optimizations are necessary to generate a lysate that more accurately reproduces the absolute *in vivo Streptomyces* transcriptionally activity.

CONCLUSION

Here, we developed a diverse library of regulatory sequences mined from BGCs to systematically characterize and dissect the transcriptional capacity of the model Actinobacteria, *S. albidoflavus* J1074 (previously *S. albus* J1074). A large range of transcriptional levels (>1000-fold) could be generated from BGC-derived regulatory sequences. To our knowledge, this data set is the largest set of regulatory sequences characterized to-date in *S. albidoflavus* using MPRAs and hopefully constitutes a useful resource for modulating gene expression and BGC activation in *S. albidoflavus*. Furthermore, we expect that transcriptional activities of BGC derived

https://doi.org/10.1021/acssynbio.0c00639
ACS Synth. Biol. XXXX, XXX, XXX−XXX

regulatory sequences we characterized can be generally extended to other evolutionarily related heterologous hosts such as *S. coelicolor*, *S. avermilitis*, and *S. lividans* to also potentially aid in heterologous BGC activations, although strain level differences in native regulatory factors may cause unexpected differences in regulatory sequence expression levels.[80,81] In this study, we identified GC content as a key sequence feature that correlated with transcriptional activity of BGC regulatory sequences. In addition, we demonstrated that the expression levels of different sets of regulatory sequences could be modulated through parallel expression of global regulatory factors including both DNA-binding transcription factors and non-DNA-binding factors. Finally, we developed an *in vitro* platform to more rapidly measure the transcriptional potential of regulatory sequences in *S. albidoflavus*. We observed transcriptional discrepancies between *in vitro* and *in vivo* measurements that appeared to be GC-dependent, which may impact future studies relying on *Streptomyces* cell-free lysates.

While we demonstrated robust transcriptional measurements of BGC regulatory sequences, one limitation is that short synthesized sequences were introduced into a heterologous host. Thus, this method would not capture any native context dependent regulatory effects, such as those of Streptomyces Antibiotic Regulatory Proteins (SARPs) which are cluster/pathway specific transcriptional regulatory proteins. Coexpression of cognate SARPs of each BGC (which can be carried out using the same expression system demonstrated here for global regulatory proteins) would enable a more focused dissection of gene regulation on a BGC specific scale. In addition, the synthesis of the length of the library also may be a limitation on our ability to capture effects of further upstream regulatory elements. We envision that further improvements in the oligonucleotide library synthesis technologies will facilitate screening of longer, more complex libraries.

Our results highlight that complex regulatory rewirings can be potentially generated *via* combinatorial expression of global regulatory proteins to further diversify gene expression patterns and assist in BGC activation. For example, the aforementioned SARPs have been shown to encode more TTA codon on average than other proteins found in the *Streptomyces* genome.[66] Therefore, coexpression of both SARPs and bldA (tRNA gene encoding the rare TTA codon) may yield more successful activation of various BGCs through synergistic interactions than what we observed with the bldA gene alone. On the other hand, coexpression of both parts, a histidine kinase and a response regulator, of a two-component system or expression of an engineered constitutively active response regulator,[82] may yield more dynamic changes in BGC expression. In addition, exploring other classes of regulatory factors, such as ECF sigma factors which are found abundantly in *Streptomyces* genomes,[83] may further increase our capacity to induce BGC activation. Lastly, improvements upon the cell-free expression system described here may yield more accurate, robust, and high-throughput systems to dissect complexity of gene regulation in BGCs. For instance, our preparation of *S. albidoflavus* cell extract employed a dialysis step that removes small molecules of less than 10 kDa. The 90-amino acid omega subunit of the *S. albidoflavus* RNA polymerase (~9.7 kDa) is nonessential but indispensable for morphological differentiation and secondary metabolism/antibiotic production in *Streptomyces*.[84−86] Further studies to assess whether the omega subunit is lost during cell-free extract preparation and any associated impact on lysate performance can provide a better mechanistically basis for our observed differences between *in vitro* and *in vivo* transcriptional capacities. In addition, the culturing regimen used to generate cell-free extract was co-opted from previous works which aimed to optimize expression yields from orthogonal transcriptional machineries. Thus, expanding upon culturing regimen, such as longer culturing time scales, to access more diverse metabolic phases may yield new cell-free extracts with more desirable transcriptional profiles. Together, our work demonstrates an application of high-throughput approaches to dissect BGC gene regulation using *S. albidoflavus* to yield deeper understanding of the transcriptional regulatory mechanisms that govern secondary metabolism.

MATERIALS AND METHODS

Mining, Synthesis, and Cloning a Regulatory Sequence Library from BGCs. From the MIBiG database (version 1.4), we selected "complete" BGC entries and mined for putative regulatory sequences. 5′ intergenic regions of every annotated coding sequences were initially selected from each BGC. Intergenic regions shorter than 100 base pairs were discarded and from the remaining sequences, 100 base pairs directly upstream of each start codons were compiled. Subsequently, sequences containing *Bam*HI or *Pst*I restriction digest cut sites were removed from the library. Finally, to each sequence, an ATG start codon and a unique 12 base pair barcode (>1 hamming distance from all other barcodes) were added to the 3′ ends. Resulting sequences were then flanked by *Bam*HI (5′) and *Pst*I (3′) restriction digest sites and common amplification sequences to yield 165 base pair sequences. Each regulatory sequence was synthesized twice with unique barcodes to yield a final oligonucleotide library of 6387 sequences (Agilent, G7721A). The oligo pool was diluted to 20 pM and amplified in 16 parallel 20 μL PCR reactions (1× Q5 Hot start HiFi Master Mix (NEB, M0543L), 0.5 μM each primer JP194, JP195) for 7 cycles to prevent overamplification (95 °C 30 s, 7 cycles: 95 °C 10 s, 72 °C 30 s; and 72 °C 2 min). PCR reactions were pooled and cleaned up with beads (Beckman Coulter, A63881). pJP50 plasmid was constructed by replacing the ermE promoter region of pIJ12057 plasmid with a synthetic DNA construct (IDT, gBlock) containing a multiple cloning site and a start codon less mCherry reporter construct flanked by *fd* and *t0* terminators using NEB Builder HiFi DNA assembly (NEB, E2621L). Resulting assembly reaction was transformed into NEB turbo competent cells (NEB, C2984H) and positive clones were selected with 50 μg/mL hygromycin (Invitrogen, 10687010). Purified oligo pool and pJP50 backbone were restriction digested with *Pst*I and *Bam*HI, PCR purified, and ligated with T4 DNA ligase (NEB M0202T). Resulting library ligation reaction was PCR purified, resuspended into a 100 μL aliquot of *E. coli* MegaX DH10B electrocompetent cells (Invitrogen, C640003), aliquoted into four prechilled 1 mm cuvettes (Bio-Rad, 1652089), electroporated, and recovered in 4 mL SOC at 37 °C for 1 h. Subsequently, 5 μL of the recovery culture were used to plate serial dilutions to determine cloning coverage (>1500× cloning coverage). Rest of the recovery culture was passaged through two subsequent liquid selections in LB with hygromycin grown overnight at 30 °C. Finally, library plasmid DNA was midiprepped (Zymo, D4200) and transformed into *E. coli S17* for subsequent conjugation reactions.

https://doi.org/10.1021/acssynbio.0c00639
ACS Synth. Biol. XXXX, XXX, XXX−XXX

Cloning Inducible Global Regulator *S. albidoflavus* Strains. Inducible regulator plasmids were generated through Gibson Assembly using NEB builder HiFi DNA assembly with PCR amplified regulator genes with overhang homology to pIJ6902 plasmid and linearized pIJ6902 plasmid backbone transformed into NEB turbo competent cells and selecting for resistant colonies with 50 μg/mL apramycin (Sigma, A2024). Resulting plasmids were miniprepped and transformed into *E. coli* S17 strain for subsequent conjugation into *S. albidoflavus*. Briefly, *E. coli* S17 strains with regulator plasmids were cultured overnight at 37 °C. Next day, the *E. coli* strains were diluted 100-fold into a 5 mL LB culture with apramycin and grown until mid log phase was reached (OD_{600} = 0.5). Then, cultures were pelleted in a prechilled centrifuge and washed twice with cold LB without antibiotics. Then, 1 μL of *S. albidoflavus* spore stock (~10^8 spores) was inoculated into 100 μL 2×YT media for each conjugation reaction and activated by heat shocking for 10 min at 50 °C. Activated spores were mixed with washed S17 strains, spun down, resuspended in 100 μL 2×YT, and plated on 30 mL ISP4 agar supplemented with 10 mM $MgCl_2$. Conjugation reactions were carried out at 30 °C for 14 h and then 1 mL antibiotic solution containing 30 μL each of 1000× apramycin and nalidixic acid (Sigma N8878) stock solution was overlaid to each conjugation plate. Selections were carried out for 14 days at 30 °C, until ex-conjugants were visible. To prep spore stock from conjugants, single colonies were inoculated into R5A media with nalidixic acid and apramycin and grown for 2 days at 30 °C. Subsequently, cells from liquid cultures were plated on ISP4 agar plates and incubated at 30 °C for 14 days then spores were harvested.

Library Scale Conjugation into *S. albidoflavus*. To transform the regulatory sequence libraries into *S. albidoflavus*, we performed an optimized conjugation protocol adapted from a previous study.[18] Briefly, we scaled up the number of S17 *E. coli* strains 10-fold (50 mL of mid log cultures) and *S. albidoflavus* spores 50-fold (~5 × 10^9 spores per conjugation) compared to a standard conjugation reaction. For each library (*i.e.*, regulator-library pairs), we performed 4 optimized conjugation reactions on 150 mm Petri dishes (Fischer, FB0875714) with 60 mL ISP4 agar with 10 mM $MgCl_2$. Selective antibiotic solution overlay consisted of 2 mL distilled water supplemented with 60 μL of 1000× hygromycin and nalidixic acid (and apramycin for regulator-library pairs) stock solutions. Following selection for 14 days at 30 °C, spores were prepped directly from the four conjugation plates to yield stocks of ~10^8 spores/μL.

Sequencing Library Preparation in *S. albidoflavus*. All *Streptomyces* cultures were propagated in 125 mL baffled Erlenmeyer flasks (KIMAX, 10−140−6) with 20−30 glass beads (Zymo, S1001) added to disrupt clumping. Initially, ~10 μL of library spore stock (~10^9 spores) were resuspended in 200 μL 2×YT media, activated through heat shock at 50 °C for 10 min and inoculated into 20 mL R5A media with nalidixic acid and hygromycin. For strains with regulators, apramycin was also included. To induced regulator expression, thiostrepton (Sigma, T8902) was supplemented to 50 μg/mL final concentration. For the BGC library in the wildtype *S. albidoflavus* strain, we cultivated 2 biological replicates and for libraries in inducible regulator *S. albidoflavus* strains, we cultivated 2 biological replicate each for induced and uninduced conditions. Following incubation in a shaker (250 rpm) at 30 °C for 48 h, cultures were diluted 50-fold into fresh 20 mL R5A media with appropriate supplements and incubated for 24 h. Dilution was repeated once more into 50 mL media, cultures were incubated for additional 16 h and then harvested. Cells from the final 50 mL cultures were pelleted through centrifugation and washed twice in cold PBS. Subsequently, the pellets were resuspended in 2.4 mL RNA-snap buffer.[87] Approximately 1.6 mL of resuspended cells were transferred into bead beating tubes (BioSpec, 330TX) preloaded with ~100 μL 0.1 mm glass beads (BioSpec, 11079101). Samples were bead beat (BioSpec, 112011) for 5 min at 4 °C in two intervals. Next, samples were subjected to heat shock at 97 °C for 7 min and spun down for 10 min at 15 000 rpm. 500 μL of supernatants were transferred to 2 mL nuclease free water in a 5 mL tube and then 2.5 mL phenol:chloroform:isoamyl alcohol (Sigma, P2069) was added. Samples were vortexed and spun down in a prechilled centrifuge at 17,500 rpm for 1 h. 2 mL of the top aqueous layer and phenol chloroform extraction was repeated once more. Next, 1.25 mL of the aqueous layer was transferred to a new 5 mL tube and ethanol precipitated (3.75 mL ethanol, 125 μL 3 M NaOAc), and resuspended in approximately 400 μL nuclease free water. 50 ng of the resulting samples were used as templates to prep DNA library sequencing samples. For each sample, regulatory sequence was amplified in a 20 μL PCR reaction: 10 μL Q5 Hot start HiFi Master Mix (NEB, M0543L), 0.2 μL 100× SYBR green (Invitrogen, S7567), 6.8 μL nuclease free water, 1 μL each of 10 mM forward and reverse primers (Supporting Information), and 1 μL template. PCR (95 °C 30 s, cycle: 95 °C 10 s, 65 °C 10 s, 72 °C 10 s; and 72 °C 2 min) was performed on a real time PCR machine (Bio-Rad, CFX-96) and the reaction was terminated during exponential amplification phase (~20 cycles). 0.1 μL of the first PCR reaction was used to perform a second 20 μL (1× Q5 Hot start HiFi Master Mix, 1× SYBR green, 0.5 μM of p5_X and p7_X amp2 primers (Supplementary Table)) PCR (95 °C 30 s, cycle: 95 °C 10 s, 72 °C 30 s; and 72 °C 2 min) reaction (ran on real time PCR machine to terminate reaction during exponential amplification) to add sample barcode indexes and Illumina p5 and p7 adapter sequences, yielding DNA library sequencing samples. For RNA libraries, we started with 50 μg of each sample and digested DNA with TURBO DNase (Invitrogen, AM2239) as follows: 2 × 50 μL reactions containing 5 μL TURBO DNase buffer, 2 μL TURBO DNase, 1 μL RiboLock RNase inhibitor (Thermo Scientific, EO0384), and 42 μL water with 25 μg RNA sample and incubated at 37 °C for 1 h. Two digestion reactions of each sample were pooled together, purified up using RNA clean and concentrator (Zymo, R1017), and eluted in 40 μL. Next, we performed three reverse transcription reactions for each sample using gene specific primer targeting the mCherry reporter transcript (JP616) and SSIV (Invitrogen, 18090010). First, 11 μL of the purified total RNA samples, 1 μL 10 mM DNTPs, and 1 μL 20 μM mCherry primer were incubated at 65 °C for 5 min and then on ice for 1 min. Next, a master mix consisting of 4 μL SSIV buffer, 1 μL 100 mM DTT, 1 μL RiboLock, and 1 μL SSIV were added and samples were incubated on a thermocycler with the following protocol: 55 °C for 90 min, 80 °C for 15 min. Next, 1 μL each of RNase A (EN0531) and RNase H (M0297S) were added and samples were incubated at 37 °C for 30 min to remove remaining RNA. Samples were purified using 2× SPRI beads (Beckman Coulter, A63881), and eluted in ~18 μL. Adapter sequences were ligated to the samples as follows: to 5.1 μL cDNA template add 2 μL 40 μM

https://doi.org/10.1021/acssynbio.0c00639
ACS Synth. Biol. XXXX, XXX, XXX−XXX

DNA adapter (JP_adapter), incubate at 75 °C for 3 min and on ice for 1 min, then add 2 μL T4 RNA ligase buffer, 0.8 μL DMSO, 0.2 μL 100 mM ATP, 8.4 μL 50% PEG 8000, and 1.5 μL T4 RNA ligase (NEB, M0437M), and incubate for 16 h at 22 °C. Adapter ligated samples were cleaned up using 2× SPRI beads and eluted in ∼20 μL. Three adapter ligation reactions were carried out for each sample. Four RNA library sequencing samples were prepared from each adapter ligated sample using the same 2 step qPCR protocol described above for prepping DNA library sequencing samples. All sequencing samples were pooled together following quantification of dsDNA concentration (Invitrogen, Q32851) of each sample, cleaned up using 2× SPRI beads (Beckman Coulter, A63881), and sequenced with an Illumina NextSeq 300-cycle (150 pair-end) midoutput kit (Illumina, 20024905). The all the raw FASTQ files for the manuscript can be accessed at ArrayExpress: E-MTAB-9223.

Sequencing Analysis and Computational Workflow. Raw sequencing reads were pair end merged using SeqPrep. Then using a custom python script, merged reads with low quality scores were removed (expected error >2 for the full merged read). Next, counts of each regulatory sequence construct with correct barcode identifiers were tallied with up to 4% mismatch tolerance in regulatory sequence regions and no mismatch allowed in barcode regions. Counts of each construct were divided by the total sum of all constructs to yield relative abundance measurements. For constructs with >10 DNA and >10 RNA counts, we calculate a Tx value by dividing its relative RNA abundance by its relative DNA abundance. Constructs with <10 DNA and <10 RNA reads were designated as missing from the library while constructs with >10 DNA reads but <10 RNA reads were designated as below the detection threshold for their transcriptional activities. Replicate Tx values (technical, biological, and barcode replicates), were averaged to yield a final set of Tx values for each sample. When Tx values between samples, each sample was median normalized prior to comparison. Sequences with more than 4-fold expression differences between induced and uninduced samples were classified as differentially expressed.

TSS and Motif Analysis. TSS calls for each RNA read was determined through alignment against the reference regulatory sequence. For each regulatory sequence with more than 200 RNA reads, we processed all TSS calls of each regulatory sequence to yield TSS locations through a k-means clustering algorithm. Briefly, starting with 16 evenly spaced seeds from 0 to 100, clusters were reduced if two clusters are found within 10 base pairs of each another or if the cluster contained <1% of all TSS calls. We defined primary TSS location of each sequence as the cluster that contained ≥70% of all TSS calls. Secondary TSS locations were defined as clusters that contained >10% of all TSS calls.

We identified enriched bipartite sequence motifs (−10 and the −35 sigma70 motifs) from regulatory sequences upstream of primary TSS locations using BioProspector.[48] BioProspector algorithm was executed with the following parameters: -d 1 -n 200 -a 1 -w 8 -W 8 -g 14 -G 19, and the regulatory sequence library was used to normalize for background distribution. Spacer between the sequence motif was determined by the median distance rounded to nearest integer value between the two motif locations across all analyzed upstream sequences.

Cloning Fluorescent Reporter Plasmid Construct. Two individual reporter plasmids were constructed to test for integration system compatibility. An eGFP insert (amplified with JP262/JP263) and pIJ10257 vector were digested with NdeI and XhoI and ligated to yield vector pJP12. An mCherry insert (amplified with JP351/352) and pIJ6902 (linearized with JP349/350) were cloned *via* Gibson Assembly through NEB builder HiFi DNA assembly.

Development and Optimization of a Cell Free Expression System Using *S. albidoflavus*. *Streptomyces albidoflavus* were propagated in 125 mL baffled flask with glass beads as described above. Compositions of all buffers and reaction mixtures can be found in Supporting Information. Briefly, 2 μL spore stock was activated in 500 μL 2×YT for 10 min at 50 °C, inoculated into 20 mL R5A media supplemented with nalidixic acid, and incubated at 30 °C for 48 h at 250 rpm. 400 μL of the resulting culture was used to inoculate fresh 20 mL of the same liquid media (50-fold dilution) and incubated for 24 h. 2 mL of the resulting culture was used to inoculate two fresh 50 mL of the same liquid media (50-fold dilution each) and incubated for 16 h. Cells were then pelleted in 50 mL falcon tubes in a prechilled centrifuge. Pellets were washed twice in 50 mL S30A buffer, and the once in 5 mL S30A buffer. Cells were then resuspended in 1 mL S30A per 1 g of wet cell mass (100 mL of cells yielded ∼6−700 mg wet cell pellet). Resuspended cells were lysed through sonication using Q125 sonicator (QSonica) while incubated on an ice slurry with the following protocol: 50% amplitude, cycles of 30 s on 30 s off until a total of 500J has been applied. Lysates were transferred to a prechilled 1.5 mL Eppendorf and centrifuged at 4 °C for 30 min at 12 000*g* to pellet residual cellular debris. Supernatant was then removed to a new Eppendorf tube and runoff reactions were performed by incubating the supernatant at 30 °C for 90 min at 250 rpm Lysates were centrifuged again for 30 min, and the supernatant was transferred to a dialysis cassette (Slide-A-Lyzer Dialysis cassette kit 10K MWCO, Thermo Scientific, 66382), and dialyzed in S30B buffer for 3 h at 4 °C. Lysates were centrifuged one last time for at 4 °C for 10 min at 12 000*g* and supernatant were removed, aliquoted, snap frozen and stored at −80 °C. Protein qubit assay kit (Invitrogen, Q33212) was used to measure protein concentration of *S. albidoflavus* cell lysate, about 18.5 mg/mL. Cell free reaction was optimized by combining the lysate, amino acids, energy buffer, PEG with varying the concentrations of Mg-glutamate (0, 3, 6, 9 mM) and K-glutamate (0, 30, 60, 90 mM) to test 16 different reaction conditions. To 7.5 μL of the reaction mix described above, we added 2 μL reporter plasmid pTOPO-F30-Broccili (final concentration of 50 mM), and 0.5 μL of 10 mM DHFBI-1T (Tocris, 5610). In negative control reactions, 2 μL of water were added instead of the reporter plasmids. Each reaction condition was tested in triplicate on a 96 well skirted PCR plate (BioRad, HSP9601) sealed with optical plate seal. Fluorescence measurements were performed in a Biotek H1 plate reader for 16 h at 30 °C using and 482 nM excitation and 505 nM emission. We observed maximum fluorescence measurements from 6 and 9 Mg-glutamate and 90 mM K-glutamate at around 5 h post incubations. We next compared fluorescence measurements using 6 and 9 mM Mg-glutamate, and 90 and 150 mM K-glutamate and observed highest fluorescence measurements from 9 mM Mg-glutamate and 150 mM K-glutamate.

Sequencing Library Prep with *S. albidoflavus* Cell Free Expression Systems. Cell free transcriptional reactions were carried out in 10 μL reactions containing 7.5 μL cell lysate and reaction mixtures and 2.5 μL plasmid DNA library (final concentration of 50 mM). Reactions were incubated at

https://doi.org/10.1021/acssynbio.0c00639
ACS Synth. Biol. XXXX, XXX, XXX−XXX

30 °C for 1 h and harvested. Reaction was purified using Zymo RCC and elution in 50 μL. 2 μL of the purified reaction was diluted 10-fold and used as a template to prep DNA sequencing libraries as described above. To prep RNA sequencing library, 5 μg of the purified reaction was used to perform a 20 μL DNase digestion reaction (2 μL Turbo DNase buffer, 1 μL Turbo DNase, 0.5 μL RiboLock RNase inhibitor). Reaction was purified using Zymo RCC kit and eluted in 20 μL. 11 μL of the resulting elution was used to perform a 20 μL reverse transcription reaction as described above, purified using 2×SPRI beads, and eluted in ∼12 μL. Two AL reactions were carried out as described above. The resulting adapter ligated samples were used to prep RNA sequencing libraries as described above.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.0c00639.

Supplementary figures on BGC meta-analysis, BGC regulatory sequence library, TSS analysis, global regulatory protein expression analysis, cell free expression analysis; List of supplementary data files; Supplementary table of global regulatory proteins, plasmids used in the study, regulatory protein sequences, oligonucleotides used in the study, statistical comparisons and associated *p*-values, and cell free expression system buffer components (PDF)

Data S1: Regulatory sequence libraries and metadata (XLSX)

Data S2: BGC Library expression data in *S. albidoflavus* (XLSX)

Data S3: BGC Library expression data with global regulatory proteins (XLSX)

Data S4: Defined regulatory sequence library (XLSX)

Data S5: BGC library expression data with *S. albidoflavus* cell free system (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Harris H. Wang − *Department of Systems Biology and Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York, New York 10032, United States;* orcid.org/0000-0003-2164-4318; Email: hw2429@columbia.edu

Authors

Jimin Park − *Department of Systems Biology and Integrated Program in Cellular, Molecular and Biomedical Studies, Columbia University Irving Medical Center, New York, New York 10032, United States;* orcid.org/0000-0001-9320-1386

Sung Sun Yim − *Department of Systems Biology and Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York, New York 10032, United States;* orcid.org/0000-0001-5060-0344

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.0c00639

Author Contributions

J.P. and H.H.W. conceived the study. J.P. performed all the experiments and data analysis with assistance from S.S.Y. and under supervision from H.H.W. Both authors wrote the paper together.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank members of the Wang lab for helpful scientific discussions and feedback. H.H.W. acknowledges relevant funding support from the NIH (U01GM110714-01A1), NSF (MCB-1453219), Sloan Foundation (FR-2015-65795), DARPA (W911NF-15-2-0065), and ONR (N00014-15-1-2704). We thank Sean Brady and members of his lab for sharing the *S. albidoflavus* J1074 strain and helpful advice on *Streptomyces* genetics and transformations.

■ ABBREVIATIONS

BGC, biosynthetic gene cluster; MPRA, massively parallel reporter assays; MIBiG, Minimum Information about a Biosynthetic Gene cluster; RS, regulatory sequences; CDS, coding sequence; KS, Kolmogorov−Smirnov; TSS, transcription start site; SARP, *Streptomyces* antibiotic regulatory proteins.

■ REFERENCES

(1) Bérdy, J. (2005) Bioactive microbial metabolites: A personal view. *J. Antibiot. 58*, 1.

(2) Vining, L. C. (1990) Functions of Secondary Metabolites. *Annu. Rev. Microbiol. 44*, 395−427.

(3) Ruiz, B., Chávez, A., Forero, A., García-Huante, Y., Romero, A., Sánchez, M., Rocha, D., Sánchez, B., Rodríguez-Sanoja, R., Sánchez, S., and Langley, E. (2010) Production of microbial secondary metabolites: Regulation by the carbon source. *Crit. Rev. Microbiol. 36*, 146−167.

(4) Jones, S. E., Pham, C. A., Zambri, M. P., McKillip, J., Carlson, E. E., and Elliot, M. A. (2019) Streptomyces volatile compounds influence exploration and microbial community dynamics by altering iron availability. *mBio 10*, 1−18.

(5) Newman, D. J., and Cragg, G. M. (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod. 75*, 311.

(6) Demain, A. L. (2014) Importance of microbial natural products and the need to revitalize their discovery. *J. Ind. Microbiol. Biotechnol. 41*, 185.

(7) Silver, L. L. (2011) Challenges of antibacterial discovery. *Clin. Microbiol. Rev. 24*, 71−109.

(8) Lok, C. (2015) Mining the microbial dark matter. *Nature 522*, 270.

(9) Rutledge, P. J., and Challis, G. L. (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol. 13*, 509.

(10) Doroghazi, J. R., Albright, J. C., Goering, A. W., Ju, K. S., Haines, R. R., Tchalukov, K. A., Labeda, D. P., Kelleher, N. L., and Metcalf, W. W. (2014) Aroadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol. 10*, 963−968.

(11) Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C., Mavrommatis, K., Pati, A., Godfrey, P. A., Koehrsen, M., Clardy, J., Birren, B. W., Takano, E., Sali, A., Linington, R. G., and Fischbach, M. A. (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell 158*, 412−421.

(12) Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C. H., Kieser, T., Larke, L., Murphy, L.,

Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J., and Hopwood, D. A. (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature 417*, 141−147.

(13) Smanski, M. J., Zhou, H., Claesen, J., Shen, B., Fischbach, M. A., and Voigt, C. A. (2016) Synthetic biology to access and expand nature's chemical diversity. *Nat. Rev. Microbiol. 14*, 135.

(14) Dashti, Y., Grkovic, T., Abdelmohsen, U., Hentschel, U., and Quinn, R. (2014) Production of Induced Secondary Metabolites by a Co-Culture of Sponge-Associated Actinomycetes, Actinokineospora sp. EG49 and Nocardiopsis sp. RV163. *Mar. Drugs 12*, 3046−3059.

(15) Abdelmohsen, U. R., Grkovic, T., Balasubramanian, S., Kamel, M. S., Quinn, R. J., and Hentschel, U. (2015) Elicitation of secondary metabolism in actinomycetes. *Biotechnol. Adv. 33*, 798.

(16) Moody, S. C. (2014) Microbial co-culture: Harnessing intermicrobial signaling for the production of novel antimicrobials. *Future Microbiol. 9*, 575.

(17) Craney, A., Ozimok, C., Pimentel-Elardo, S. M., Capretta, A., and Nodwell, J. R. (2012) Chemical perturbation of secondary metabolism demonstrates important links to primary metabolism. *Chem. Biol. 19*, 1020−1027.

(18) Iqbal, H. A., Low-Beinart, L., Obiajulu, J. U., and Brady, S. F. (2016) Natural Product Discovery through Improved Functional Metagenomics in Streptomyces. *J. Am. Chem. Soc. 138*, 9341−4.

(19) Katz, M., Hover, B. M., and Brady, S. F. (2016) Culture-independent discovery of natural products from soil metagenomes. *J. Ind. Microbiol. Biotechnol. 43*, 129−141.

(20) Brady, S. F. (2007) Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nat. Protoc. 2*, 1297−1305.

(21) Banik, J. J., and Brady, S. F. (2008) Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. *Proc. Natl. Acad. Sci. U. S. A. 105*, 17273−17277.

(22) Kang, H. S., Charlop-Powers, Z., and Brady, S. F. (2016) Multiplexed CRISPR/Cas9- and TAR-Mediated Promoter Engineering of Natural Product Biosynthetic Gene Clusters in Yeast. *ACS Synth. Biol. 5*, 1002−1010.

(23) Shao, Z., Rao, G., Li, C., Abil, Z., Luo, Y., and Zhao, H. (2013) Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. *ACS Synth. Biol. 2*, 662−669.

(24) Montiel, D., Kang, H. S., Chang, F. Y., Charlop-Powers, Z., and Brady, S. F. (2015) Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. *Proc. Natl. Acad. Sci. U. S. A. 112*, 8953−8958.

(25) Elena, C., Ravasi, P., Castelli, M. E., Peirú, S., and Menzella, H. G. (2014) Expression of codon optimized genes in microbial systems: Current industrial applications and perspectives. *Front. Microbiol.*, DOI: 10.3389/fmicb.2014.00021.

(26) Park, J., and Wang, H. H. (2018) Systematic and synthetic approaches to rewire regulatory networks. *Current Opinion in Systems Biology. 8*, 90.

(27) Chen, Y., Wendt-Pienkowski, E., and Shen, B. (2008) Identification and Utility of FdmR1 as a Streptomyces Antibiotic Regulatory Protein Activator for Fredericamycin Production in Streptomyces griseus ATCC 49344 and Heterologous Hosts. *J. Bacteriol. 190*, 5587−5596.

(28) Olano, C., García, I., González, A., Rodriguez, M., Rozas, D., Rubio, J., Sánchez-Hidalgo, M., Braña, A. F., Méndez, C., and Salas, J. A. (2014) Activation and identification of five clusters for secondary metabolites in Streptomyces albusJ1074. *Microb. Biotechnol. 7*, 242−256.

(29) Tanaka, Y., Kasahara, K., Hirose, Y., Murakami, K., Kugimiya, R., and Ochi, K. (2013) Activation and Products of the Cryptic Secondary Metabolite Biosynthetic Gene Clusters by Rifampin Resistance (rpoB) Mutations in Actinomycetes. *J. Bacteriol. 195*, 2959−2970.

(30) Hosaka, T., Ohnishi-Kameyama, M., Muramatsu, H., Murakami, K., Tsurumi, Y., Kodani, S., Yoshida, M., Fujie, A., and Ochi, K. (2009) Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nat. Biotechnol. 27*, 462−464.

(31) Johns, N. I., Gomes, A. L. C., Yim, S. S., Yang, A., Blazejewski, T., Smillie, C. S., Smith, M. B., Alm, E. J., Kosuri, S., and Wang, H. H. (2018) Metagenomic mining of regulatory elements enables programmable species-selective gene expression. *Nat. Methods 15*, 323−329.

(32) Kosuri, S., Goodman, D. B., Cambray, G., Mutalik, V. K., Gao, Y., Arkin, A. P., Endy, D., and Church, G. M. (2013) Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A. 110*, 14024−14029.

(33) Cambray, G., Guimaraes, J. C., and Arkin, A. P. (2018) Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in escherichia coli. *Nat. Biotechnol. 36*, 1005.

(34) Labeda, D. P., Doroghazi, J. R., Ju, K. S., and Metcalf, W. W. (2014) Taxonomic evaluation of Streptomyces albus and related species using multilocus sequence analysis and proposals to emend the description of Streptomyces albus and describe Streptomyces pathocidini sp. nov. *Int. J. Syst. Evol. Microbiol. 64*, 894−900.

(35) Kallifidas, D., Jiang, G., Ding, Y., and Luesch, H. (2018) Rational engineering of Streptomyces albus J1074 for the overexpression of secondary metabolite gene clusters. *Microb. Cell Fact. 17*, 25.

(36) Zaburannyi, N., Rabyk, M., Ostash, B., Fedorenko, V., and Luzhetskyy, A. (2014) Insights into naturally minimised Streptomyces albus J1074 genome. *BMC Genomics 15*, 97.

(37) Ji, C.-H., Kim, J.-P., and Kang, H.-S. (2018) Library of Synthetic Streptomyces Regulatory Sequences for Use in Promoter Engineering of Natural Product Biosynthetic Gene Clusters. *ACS Synth. Biol. 7*, 1946.

(38) Luo, Y., Zhang, L., Barton, K. W., and Zhao, H. (2015) Systematic Identification of a Panel of Strong Constitutive Promoters from Streptomyces albus. *ACS Synth. Biol. 4*, 1001−10.

(39) Sohoni, S. V., Fazio, A., Workman, C. T., Mijakovic, I., and Lantz, A. E. (2014) Synthetic Promoter Library for Modulation of Actinorhodin Production in Streptomyces coelicolor A3(2). *PLoS One 9*, e99701.

(40) Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., de Bruijn, I., Chooi, Y. H., Claesen, J., Coates, R. C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D. J., Fewer, D. P., Garg, N., Geiger, C., Gomez-Escribano, J. P., Greule, A., Hadjithomas, M., Haines, A. S., Helfrich, E. J. N., Hillwig, M. L., Ishida, K., Jones, A. C., Jones, C. S., Jungmann, K., Kegler, C., Kim, H. U., Kötter, P., Krug, D., Masschelein, J., Melnik, A. v., Mantovani, S. M., Monroe, E. A., Moore, M., Moss, N., Nützmann, H. W., Pan, G., Pati, A., Petras, D., Reen, F. J., Rosconi, F., Rui, Z., Tian, Z., Tobias, N. J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X., Yim, G., Yu, F., Xie, Y., Aigle, B., Apel, A. K., Balibar, C. J., Balskus, E. P., Barona-Gómez, F., Bechthold, A., Bode, H. B., Borriss, R., Brady, S. F., Brakhage, A. A., Caffrey, P., Cheng, Y. Q., Clardy, J., Cox, R. J., de Mot, R., Donadio, S., Donia, M. S., van der Donk, W. A., Dorrestein, P. C., Doyle, S., Driessen, A. J. M., Ehling-Schulz, M., Entian, K. D., Fischbach, M. A., Gerwick, L., Gerwick, W. H., Gross, H., Gust, B., Hertweck, C., Höfte, M., Jensen, S. E., Ju, J., Katz, L., Kaysser, L., Klassen, J. L., Keller, N. P., Kormanec, J., Kuipers, O. P., Kuzuyama, T., Kyrpides, N. C., Kwon, H. J., Lautru, S., Lavigne, R., Lee, C. Y., Linquan, B., Liu, X., Liu, W., Luzhetskyy, A., Mahmud, T., Mast, Y., Méndez, C., Metsä-Ketelä, M., Micklefield, J., Mitchell, D. A., Moore, B. S., Moreira, L. M., Müller, R., Neilan, B. A., Nett, M., Nielsen, J., O'Gara, F., Oikawa, H., Osbourn, A., Osburne, M. S., Ostash, B., Payne, S. M., Pernodet, J. L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J. M., Salas, J. A., Schmitt, E. K., Scott, B., Seipke, R. F., Shen, B., Sherman, D. H., Sivonen, K., Smanski, M. J., Sosio, M., Stegmann, E., Süssmuth, R. D., Tahlan, K., Thomas, C. M., Tang, Y., Truman, A. W., Viaud, M., Walton, J. D., Walsh, C. T., Weber, T., van Wezel, G. P., Wilkinson, B., Willey, J. M., Wohlleben, W., Wright, G.

https://doi.org/10.1021/acssynbio.0c00639
ACS Synth. Biol. XXXX, XXX, XXX−XXX

D., Ziemert, N., Zhang, C., Zotchev, S. B., Breitling, R., Takano, E., and Glöckner, F. O. (2015) Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.*, 625.
(41) Gomes, A. L. C., Johns, N. I., Yang, A., Velez-Cortes, F., Smillie, C. S., Smith, M. B., Alm, E. J., and Wang, H. H. (2020) Genome and sequence determinants governing the expression of horizontally acquired DNA in bacteria. *ISME J. 14*, 2347.
(42) Ruiz, B., Chávez, A., Forero, A., García-Huante, Y., Romero, A., Snchez, M., Rocha, D., Snchez, B., Rodríguez-Sanoja, R., Sánchez, S., and Langley, E. (2010) Production of microbial secondary metabolites: Regulation by the carbon source. *Crit. Rev. Microbiol. 36*, 146.
(43) Murakami, T., Holt, T. G., and Thompson, C. J. (1989) Thiostrepton-induced gene expression in Streptomyces lividans. *J. Bacteriol. 171*, 1459−1466.
(44) Bibb, M. J., Janssen, G. R., and Ward, J. M. (1985) Cloning and analysis of the promoter region of the erythromycin resistance gene (ermE) of Streptomyces erythraeus. *Gene 38*, 215−226.
(45) Hong, H. J., Hutchings, M. I., Hill, L. M., and Buttner, M. J. (2005) The role of the novel fem protein VanK in vancomycin resistance in Streptomyces coelicolor. *J. Biol. Chem. 280*, 13055−13061.
(46) Yim, S. S., Johns, N. I., Park, J., Gomes, A. L., McBee, R. M., Richardson, M., Ronda, C., Chen, S. P., Garenne, D., Noireaux, V., and Wang, H. H. (2019) Multiplex transcriptional characterizations across diverse bacterial species using cell-free systems. *Mol. Syst. Biol.*, DOI: 10.15252/msb.20198875.
(47) Jeong, Y., Kim, J. N., Kim, M. W., Bucca, G., Cho, S., Yoon, Y. J., Kim, B. G., Roe, J. H., Kim, S. C., Smith, C. P., and Cho, B. K. (2016) The dynamic transcriptional and translational landscape of the model antibiotic producer Streptomyces coelicolor A3(2). *Nat. Commun. 7*, 1−11.
(48) Liu, X., Brutlag, D. L., and Liu, J. S. (2001) BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing 138*, 127−138.
(49) Šmídová, K., Ziková, A., Pospíšil, J., Schwarz, M., Bobek, J., and Vohradsky, J. (2019) DNA mapping and kinetic modeling of the HrdB regulon in Streptomyces coelicolor. *Nucleic Acids Res. 47*, 621−633.
(50) Mejía-Almonte, C., Busby, S. J. W., Wade, J. T., van Helden, J., Arkin, A. P., Stormo, G. D., Eilbeck, K., Palsson, B. O., Galagan, J. E., and Collado-Vides, J. (2020) Redefining fundamental concepts of transcription initiation in bacteria. *Nat. Rev. Genet. 21*, 699.
(51) Urtecho, G., Tripp, A. D., Insigne, K. D., Kim, H., and Kosuri, S. (2019) Systematic Dissection of Sequence Elements Controlling σ70 Promoters Using a Genomically Encoded Multiplexed Reporter Assay in *Escherichia coli*. *Biochemistry 58*, 1539−1551.
(52) Horbal, L., Fedorenko, V., Bechthold, A., and Luzhetskyy, A. (2013) A transposon-based strategy to identify the regulatory gene network responsible for landomycin E biosynthesis. *FEMS Microbiol. Lett. 342*, 138−146.
(53) Ostash, B., Rebets, Y., Myronovskyy, M., Tsypik, O., Ostash, I., Kulachkovskyy, O., Datsyuk, Y., Nakamura, T., Walker, S., and Fedorenko, V. (2011) Identification and characterization of the Streptomyces globisporus 1912 regulatory gene lndYR that affects sporulation and antibiotic production. *Microbiology 157*, 1240−1249.
(54) Daniel-Ivad, M., Hameed, N., Tan, S., Dhanjal, R., Socko, D., Pak, P., Gverzdys, T., Elliot, M. A., and Nodwell, J. R. (2017) An Engineered Allele of afsQ1 Facilitates the Discovery and Investigation of Cryptic Natural Products. *ACS Chem. Biol. 12*, 628−634.
(55) Rico, S., Santamaría, R. I., Yepes, A., Rodríguez, H., Laing, E., Bucca, G., Smith, C. P., and Díaz, M. (2014) Deciphering the regulon of Streptomyces coelicolor AbrC3, a positive response regulator of antibiotic production. *Appl. Environ. Microbiol. 80*, 2417−2428.
(56) RodrÃguez, H., Rico, S., Yepes, A., Franco-EchevarrÃa, E., Antoraz, S., SantamarÃa, R. I., and DÃaz, M. (2015) The two kinases, AbrC1 and AbrC2, of the atypical two-component system AbrC are needed to regulate antibiotic production and differentiation in Streptomyces coelicolor. *Front. Microbiol. 6*, 450.
(57) Becerril, A., Álvarez, S., Braña, A. F., Rico, S., Díaz, M., Santamaría, R. I., Salas, J. A., and Méndez, C. (2018) Uncovering production of specialized metabolites by Streptomyces argillaceus: Activation of cryptic biosynthesis gene clusters using nutritional and genetic approaches. *PLoS ONE 13*, e0198145.
(58) Ishizuka, H., Horinouchi, S., Kieser, H. M., Hopwood, D. A., and Beppu, T. (1992) A putative two-component regulatory system involved in secondary metabolism in Streptomyces spp. *J. Bacteriol. 174*, 7585−7594.
(59) Wang, R., Mast, Y., Wang, J., Zhang, W., Zhao, G., Wohlleben, W., Lu, Y., and Jiang, W. (2013) Identification of two-component system AfsQ1/Q2 regulon and its cross-regulation with GlnR in Streptomyces coelicolor. *Mol. Microbiol. 87*, 30−48.
(60) Chen, S., Zheng, G., Zhu, H., He, H., Chen, L., Zhang, W., Jiang, W., and Lu, Y. (2016) Roles of two-component system AfsQ1/Q2 in regulating biosynthesis of the yellow-pigmented coelimycin P2 in Streptomyces coelicolor. *FEMS Microbiol. Lett. 363*, fnw160.
(61) Shu, D., Chen, L., Wang, W., Yu, Z., Ren, C., Zhang, W., Yang, S., Lu, Y., and Jiang, W. (2009) afsQ1-Q2-sigQ is a pleiotropic but conditionally required signal transduction system for both secondary metabolism and morphological development in Streptomyces coelicolor. *Appl. Microbiol. Biotechnol. 81*, 1149−1160.
(62) Hutchings, M. I., Hong, H. J., Leibovitz, E., Sutcliffe, I. C., and Buttner, M. J. (2006) The σE cell envelope stress response of Streptomyces coelicolor is influenced by a novel lipoprotein, CseA. *J. Bacteriol. 188*, 7222−7229.
(63) Jones, G. H., Paget, M. S. B., Chamberlin, L., and Buttner, M. J. (1997) Sigma-E is required for the production of the antibiotic actinomycin in Streptomyces antibioticus. *Mol. Microbiol. 23*, 169−178.
(64) Hutchings, M. I., Hong, H. J., Leibovitz, E., Sutcliffe, I. C., and Buttner, M. J. (2006) The σE cell envelope stress response of Streptomyces coelicolor is influenced by a novel lipoprotein, CseA. *J. Bacteriol. 188*, 7222−7229.
(65) Hesketh, A., Hill, C., Mokhtar, J., Novotna, G., Tran, N., Bibb, M., and Hong, H. J. (2011) Genome-wide dynamics of a bacterial response to antibiotics that target the cell envelope. *BMC Genomics 12*, 226.
(66) Leskiw, B. K., Lawlor, E. J., Fernandez-Abalos, J. M., and Chater, K. F. (1991) TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative, Streptomyces mutants. *Proc. Natl. Acad. Sci. U. S. A. 88*, 2461−2465.
(67) Hackl, S., and Bechthold, A. (2015) The Gene bldA, a Regulator of Morphological Differentiation and Antibiotic Production in Streptomyces. *Arch. Pharm.*, DOI: 10.1002/ardp.201500073.
(68) Pope, M. K., Green, B., and Westpheling, J. (1998) The bldB gene encodes a small protein required for morphogenesis, antibiotic production, and catabolite control in Streptomyces coelicolor. *J. Bacteriol. 180*, 1556−1562.
(69) Eccleston, M., Ali, R. A., Seyler, R., Westpheling, J., and Nodwell, J. (2002) Structural and genetic analysis of the BldB protein of Streptomyces coelicolor. *J. Bacteriol. 184*, 4270−4276.
(70) Liu, G., Chater, K. F., Chandra, G., Niu, G., and Tan, H. (2013) Molecular Regulation of Antibiotic Biosynthesis in Streptomyces. *Microbiol. Mol. Biol. Rev. 77*, 112−143.
(71) Hosaka, T., Xu, J., and Ochi, K. (2006) Increased expression of ribosome recycling factor is responsible for the enhanced protein synthesis during the late growth phase in an antibiotic-overproducing Streptomyces coelicolor ribosomal rpsL mutant. *Mol. Microbiol. 61*, 883−897.
(72) Li, L., Guo, J., Wen, Y., Chen, Z., Song, Y., and Li, J. (2010) Overexpression of ribosome recycling factor causes increased production of avermectin in Streptomyces avermitilis strains. *J. Ind. Microbiol. Biotechnol. 37*, 673−679.
(73) Ma, Z., Tao, L., Bechthold, A., Shentu, X., Bian, Y., and Yu, X. (2014) Overexpression of ribosome recycling factor is responsible for improvement of nucleotide antibiotic-toyocamycin in Streptomyces

https://doi.org/10.1021/acssynbio.0c00639
ACS Synth. Biol. XXXX, XXX, XXX−XXX

diastatochromogenes 1628. *Appl. Microbiol. Biotechnol. 98*, 5051−5058.
(74) Gregory, M. A., Till, R., and Smith, M. C. M. (2003) Integration site for Streptomyces phage φBT1 and development of site-specific integrating vectors. *J. Bacteriol. 185*, 5320−5323.
(75) Silverman, A. D., Karim, A. S., and Jewett, M. C. (2020) Cell-free gene expression: an expanded repertoire of applications. *Nat. Rev. Genet. 21*, 151.
(76) Li, J., Wang, H., Kwon, Y.-C., and Jewett, M. C. (2017) Establishing a high yielding streptomyces -based cell-free protein synthesis system. *Biotechnol. Bioeng. 114*, 1343−1353.
(77) Moore, S. J., Lai, H.-E., Needham, H., Polizzi, K. M., and Freemont, P. S. (2017) *Streptomyces venezuelae* TX-TL - a next generation cell-free synthetic biology tool. *Biotechnol. J. 12*, 1600678.
(78) Li, J., Wang, H., and Jewett, M. C. (2018) Expanding the palette of Streptomyces-based cell-free protein synthesis systems with enhanced yields. *Biochem. Eng. J. 130*, 29−33.
(79) Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., and Noireaux, V. (2013) Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. *J. Visualized Exp.*, No. e50762.
(80) Hsiao, N., and Kirby, R. (2008) Comparative genomics of Streptomyces avermitilis, Streptomyces cattleya, Streptomyces maritimus and Kitasatospora aureofaciens using a Streptomyces coelicolor microarray system. *Antonie van Leeuwenhoek 93*, 1−25.
(81) Lee, Y., Lee, N., Jeong, Y., Hwang, S., Kim, W., Cho, S., Palsson, B. O., and Cho, B.-K. (2019) The Transcription Unit Architecture of Streptomyces lividans TK24. *Front. Microbiol. 10*, 2074.
(82) Daniel-Ivad, M., Hameed, N., Tan, S., Dhanjal, R., Socko, D., Pak, P., Gverzdys, T., Elliot, M. A., and Nodwell, J. R. (2017) An Engineered Allele of afsQ1 Facilitates the Discovery and Investigation of Cryptic Natural Products. *ACS Chem. Biol. 12*, 628−634.
(83) Paget, M. (2015) Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules 5*, 1245−1265.
(84) Kojima, I., Kasuga, K., Kobayashi, M., Fukasawa, A., Mizuno, S., Arisawa, A., and Akagawa, H. (2002) The rpoZ gene, encoding the RNA polymerase omega subunit, is required for antibiotic production and morphological differentiation in Streptomyces kasugaensis. *J. Bacteriol. 184*, 6417−6423.
(85) Santos-Beneit, F., Barriuso-Iglesias, M., Fernández-Martínez, L. T., Martínez-Castro, M., Sola-Landa, A., Rodríguez-García, A., and Martín, J. F. (2011) The RNA polymerase omega factor rpoZ is regulated by phop and has an important role in antibiotic biosynthesis and morphological differentiation in Streptomyces coelicolor. *Appl. Environ. Microbiol. 77*, 7586−7594.
(86) Tran, N. T., Huang, X., Hong, H. J., Bush, M. J., Chandra, G., Pinto, D., Bibb, M. J., Hutchings, M. I., Mascher, T., and Buttner, M. J. (2019) Defining the regulon of genes controlled by σE, a key regulator of the cell envelope stress response in Streptomyces coelicolor. *Mol. Microbiol. 112*, 461−481.
(87) Stead, M. B., Agrawal, A., Bowden, K. E., Nasir, R., Mohanty, B. K., Meagher, R. B., and Kushner, S. R. (2012) RNAsnap: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria. *Nucleic Acids Res. 40*, No. e156.

https://doi.org/10.1021/acssynbio.0c00639
ACS Synth. Biol. XXXX, XXX, XXX−XXX