

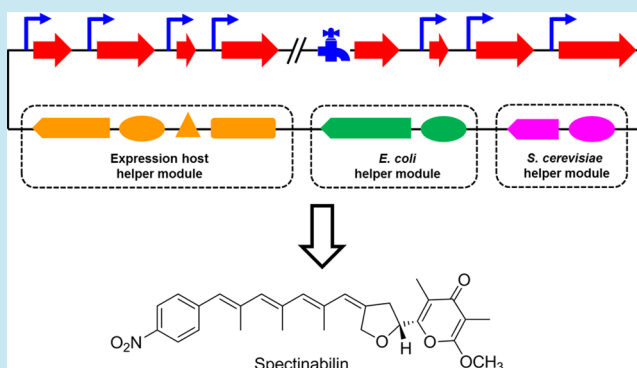
# Refactoring the Silent Spectinabilin Gene Cluster Using a Plug-and-Play Scaffold

Zengyi Shao,<sup>†,||</sup> Guodong Rao,<sup>‡</sup> Chun Li,<sup>†</sup> Zhanar Abil,<sup>§</sup> Yunzi Luo,<sup>†</sup> and Huimin Zhao<sup>\*,†,‡,§</sup><sup>†</sup>Department of Chemical and Biomolecular Engineering, <sup>‡</sup>Department of Chemistry, and <sup>§</sup>Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States

## Supporting Information

**ABSTRACT:** Natural products (secondary metabolites) are a rich source of compounds with important biological activities. Eliciting pathway expression is always challenging but extremely important in natural product discovery because an individual pathway is tightly controlled through a unique regulation mechanism and hence often remains silent under the routine culturing conditions. To overcome the drawbacks of the traditional approaches that lack general applicability, we developed a simple synthetic biology approach that decouples pathway expression from complex native regulations. Briefly, the entire silent biosynthetic pathway is refactored using a plug-and-play scaffold and a set of heterologous promoters that are functional in a heterologous host under the target culturing condition. Using this strategy, we successfully awakened the silent spectinabilin pathway from *Streptomyces orinoci*. This strategy bypasses the traditional laborious processes to elicit pathway expression and represents a new platform for discovering novel natural products.

**KEYWORDS:** natural products, silent pathways, genome mining, plug-and-play scaffold, synthetic biology, pathway assembly



Microorganisms have evolved to produce a myriad array of complex molecules known as natural products or secondary metabolites, many of which possess important biological activities such as antibacterial, antiviral, and anticancer properties.<sup>1,2</sup> The rapidly increasing number of sequenced genomes and metagenomes provide a tremendously rich source for the discovery of gene clusters involved in the biosynthesis of new compounds. However, the discovery and economical production of natural products are hampered by our limited ability to manipulate most organisms, determine suitable conditions for eliciting pathway expression and produce sufficient amount of products for structure elucidation.

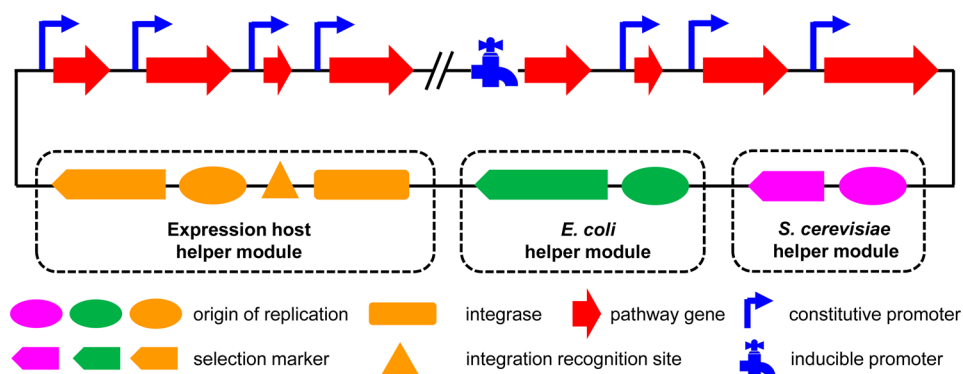
The biosynthesis of natural products is highly regulated, and gene clusters often remain silent until suitable conditions are met. The regulation is conducted through dozens of pleiotropic regulatory genes and pathway-specific regulators.<sup>3–6</sup> They interact with each other to form an extremely complex network in response to a variety of physiological and environmental signals. Existing approaches to studying natural product biosynthetic clusters mainly include (i) manipulating cell culture parameters, such as medium composition, to ensure expression of pathway-specific activator(s), the presence of physiological and environmental co-inducers, or derepression of the genes repressed by repressor(s);<sup>4,6</sup> (ii) engineering the regulation by expressing the pathway-specific regulator under a well-characterized promoter;<sup>7,8</sup> (iii) testing a variety of heterologous hosts to express the target cluster;<sup>9</sup> (iv) silencing

major secondary metabolite biosynthetic pathways to simplify product identification and relieve competition for key precursors;<sup>10</sup> and (v) utilizing industrial strains that have been set up for high-level production of specific compounds.<sup>11</sup> All these strategies can only be applied on a case-by-case basis. Each gene cluster has its own unique regulatory mechanism that must be examined individually to identify the suitable context for the cluster to be activated. Our current understanding of regulation hierarchy is insufficient to accurately predict the functions of all the regulatory elements. Thus, it is highly desirable to develop generally applicable approaches that reduce regulation complexity without any requirement of specific tailoring to the cryptic biosynthetic pathway of interest.

Here we describe a synthetic biology-based strategy for decoupling pathway expression from native sophisticated regulation cascades. On the basis of the key engineering principle of modular design in synthetic biology, the scaffold consists of three types of modules (Scheme 1): promoter modules, gene modules, and helper modules. For promoter modules, strong promoters confirmed in the target expression host under the selected culture condition are chosen to ensure the transcription of downstream genes. They can be cloned from the potential promoter regions of the endogenous housekeeping genes in the target expression host. In addition,

Received: May 18, 2013

Published: August 7, 2013

Scheme 1. Design of a Plug-and-Play Scaffold for Refactoring Natural Product Biosynthetic Pathways<sup>a</sup>

<sup>a</sup>The scaffold consists of promoter modules, gene modules, and helper modules. The refactoring strategy is to select a single heterologous host, identify a set of strong promoters under a target culture condition, assemble individual biosynthetic genes with these promoters into a new gene cluster, and express the refactored gene cluster in the heterologous host under the target culture condition.

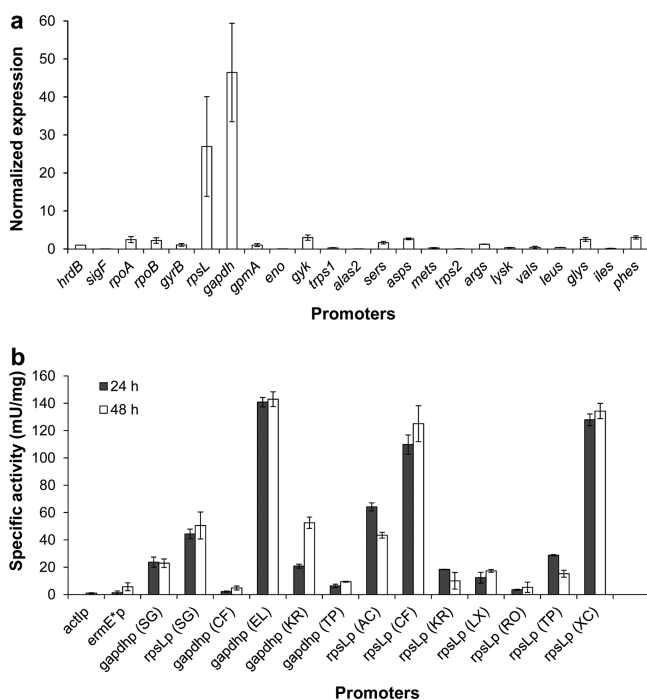
other organisms closely related to the target expression host can also be used to identify strong promoters that can be recognized by the target expression host. For helper modules, genetic elements (mainly including an origin of replication and a selection marker) needed for DNA maintenance and replication in individual hosts are amplified from the corresponding vectors. Typically, hosts include the cluster assembly host *Saccharomyces cerevisiae*, the DNA enrichment host *Escherichia coli*, and the target expression host. After promoter modules and helper modules are set up, genes and their downstream intergenic sequences in the target cluster can be individually amplified from the isolated genomic DNA if the native host can be cultivated or obtained directly via chemical synthesis and subsequently plugged into each gene module. The downstream intergenic sequences are included because some of them might contain terminator function (see more discussion below). In some cases, if an inducible promoter is available for the target expression host, it can be placed upstream of the gene encoding the enzyme catalyzing the first step in the biosynthesis. The necessity for including such a promoter whose activity is controlled by an exogenously added inducer depends on whether the toxicity of the final product or the biosynthetic intermediates to the expression host is a concern. The assembly of such an artificial gene cluster is based on the DNA assembler approach that relies on yeast homologous recombination to splice multiple overlapping DNA fragments.<sup>12,13</sup> Using such a plug-and-play scaffold, the sophisticated regulation embedded in individual clusters is removed and replaced with a set of regulations that is predictable, easy to manipulate, and not specifically linked to any gene cluster. Such a strategy offers a new platform for *de novo* cluster assembly and genome mining for discovering new natural products.

## RESULTS AND DISCUSSION

Because a full set of constitutive or inducible promoters needed for the plug-and-play strategy mentioned above are typically not available for most expression hosts except *E. coli* and *S. cerevisiae*, it is important to quickly discover as many strong promoters as possible for a target expression host. We chose to work on gene clusters from Streptomycetes because they are prolific sources of bioactive natural products, many of which exhibit important medical properties.<sup>14,15</sup> *Streptomyces lividans*, a laboratory strain that has been used extensively for studying

gene clusters from Streptomycetes, was chosen as a target heterologous host because of its high conjugative DNA transfer efficiency and the availability of several genetic manipulation tools.<sup>9,16–20</sup>

We first selected promoter candidates upstream of 23 housekeeping genes originated from *Streptomyces griseus*, whose genome sequence is available. These genes include RNA polymerase subunits, elongation factors, ribosomal proteins, glycolytic enzymes, and various amino-acyl tRNA synthetases (Table S1 of the Supporting Information). Using real-time polymerase chain reaction (PCR), two genes encoding glyceraldehyde-3-phosphate dehydrogenase (*gapdh*) and 30S ribosomal protein S12 (*rpsL*), stood out with transcription levels much higher than those of all the other genes under our fixed culturing condition (Figure 1a), indicating that their corresponding promoters could be very strong. The *gapdh* promoter, named *gapdh<sub>p</sub>* (SG), is located upstream of the *gapdh* operon consisting of *Gapdh*, phosphoglycerate kinase (*pgk*), and triosephosphate isomerase (*tpiA*), the enzymes catalyzing the sixth, seventh, and fifth steps, respectively, in the glycolysis pathway; the *rpsL* promoter, named *rpsL<sub>p</sub>* (SG), resides upstream of another operon consisting of 30s ribosomal proteins S12 and S7, and elongation factor G and Tu (Figure S1 of the Supporting Information). To ensure that these promoters can be used to drive the transcription of heterologous genes and also compare their activities with those reported in the literature, the intergenic region between the gene located upstream and the *gapdh* operon (or the *rpsL* operon) was subsequently fused with the Streptomyces reporter gene, catechol 2,3-dioxygenase (*xylE*), which catalyzes the conversion of colorless catechol to yellow-colored 2-hydroxymuconic semialdehyde. The *xylE* activity assay<sup>20</sup> confirmed that their activities are much stronger than the control promoter, *ermE\*<sub>p</sub>* (Figure 1b), which is the mutated variant of the promoter of the erythromycin resistance gene from *Saccharopolyspora erythraea* and is believed to be one of the strongest constitutive promoters in Streptomycetes.<sup>21,22</sup> The strong activities of *gapdh<sub>p</sub>* and the promoters of various translation elongation factors have also been observed in many other microorganisms, such as fungi, bacteria, microalgae, and protozoa.<sup>23–29</sup> Encouraged by the strong activities of *gapdh<sub>p</sub>* (SG) and *rpsL<sub>p</sub>* (SG), we decided to examine their equivalents from other species. The promoters from different *Streptomyces* species share extremely high degrees of homology with each

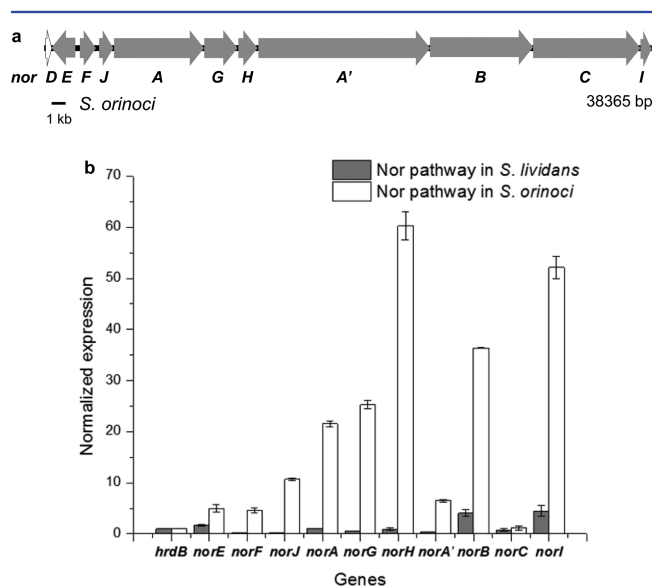


**Figure 1.** Promoter screening for pathway refactoring in *Streptomyces*. (a) Identification of strong constitutive promoters by real-time PCR analysis of the transcription of 23 housekeeping genes in *S. griseus*. Samples were taken at different time points. The *y*-axis scale represents the expression value relative to that of *HrdB*, a commonly used housekeeping sigma factor,<sup>56–58</sup> which was set to 1. The levels of expression of *gapdh* and *rpsL* were higher than those of the other genes at all the sampling points, with the 12 h samples showing the highest level of significance (46- and 27-fold higher than that of *hrdB*, respectively). (b) Evaluation of the activities of the heterologous promoters using *xylE* as a reporter. The entire intergenic region between the target gene and its upstream gene was cloned upstream of *xylE*. Here we did not experimentally determine the ribosomal binding site (RBS) for each promoter and assumed it is located 6–10 bp upstream of each start codon. However, we did find these intergenic regions are AG-rich for most promoters. Promoters act<sub>1</sub>P and ermE\*<sub>p</sub> are the two commonly used promoters reported in the literature.<sup>21,22,59</sup> The two letters in parentheses represent the names of individual actinomycetes (Table S2 of the Supporting Information).

other (Figure S2a of the Supporting Information), which is undesired because such high degrees of sequence similarity would cause severe deletions during DNA assembly in *S. cerevisiae*. Because *Streptomyces* belong to the family of actinobacteria, the promoters from other genera of actinobacteria could be possibly recognized by the transcription machinery in *Streptomyces* as well. The *gapdh* and the *rpsL* operon were found to be highly conserved in the family of actinobacteria, but the corresponding promoters are highly diversified (Figure S2b of the Supporting Information). Next, the potential *gapdh*<sub>p</sub> and *rpsL*<sub>p</sub> from 18 distinct actinobacteria (Figure S3 of the Supporting Information) were cloned upstream of *xylE* for more quantitative comparison. As a result, 13 of the 36 promoter candidates were shown to be very active in *S. lividans*, many of them having more than 10-fold higher activities than ermE\*<sub>p</sub> (Figure 1b and Table S2 of the Supporting Information).

As proof of concept, the spectinabilin gene cluster from *Streptomyces orinoci*<sup>30</sup> was chosen as a model pathway. Spectinabilin, isolated from *Streptomyces spectabilis* and *S.*

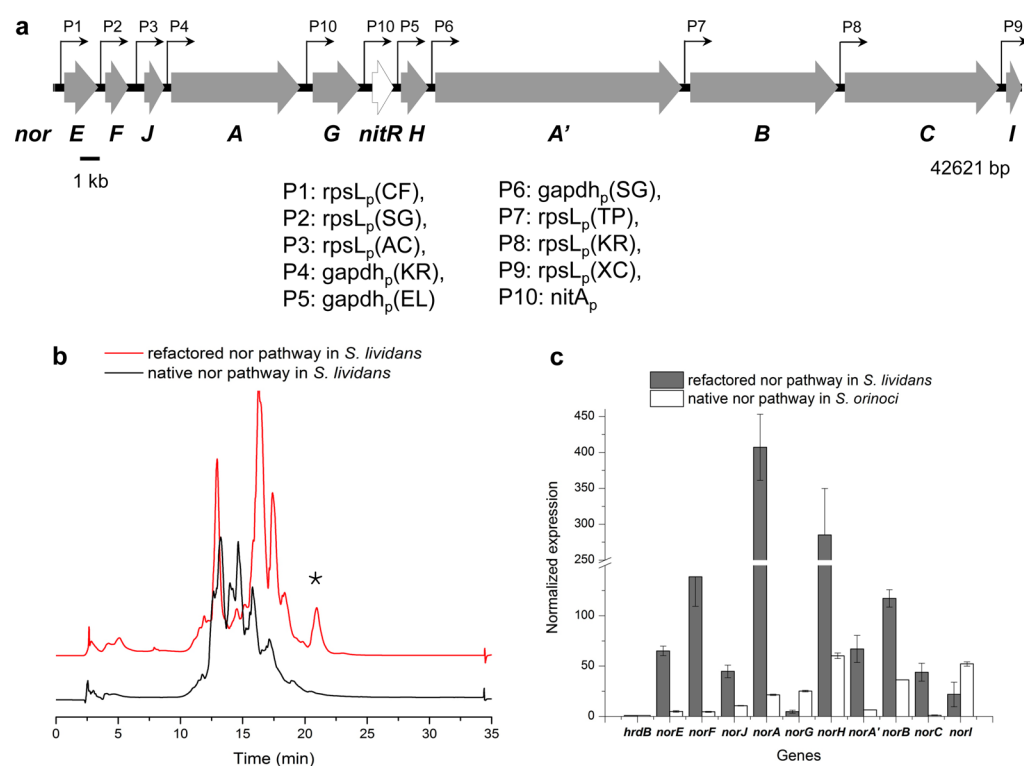
*orinoci*, is a nitrophenyl-substituted polyketide exhibiting antimalarial and antiviral activities.<sup>31,32</sup> Interestingly, although the catalytic proteins from the two spectinabilin clusters share a very high degree of homology, these two clusters undergo different regulations. The cluster from *S. spectabilis* (named the *spn* cluster) contains SpnD as an activator, and the cluster from *S. orinoci* (named the *nor* cluster) contains NorD as a repressor (Figure 2a and Figure S4 of the Supporting Information).<sup>33</sup>



**Figure 2.** (a) Spectinabilin gene cluster from *S. orinoci*. (b) Real-time PCR analysis of the *nor* gene transcription in *S. lividans* and *S. orinoci*.

When the two clusters were cloned into *S. lividans*, only the *spn* cluster could heterologously produce spectinabilin under the laboratory conditions.<sup>33,34</sup> We first reconstructed the *nor* cluster excluding *norD* in *S. lividans*. However, no production of spectinabilin was observed either (Figure 3b), indicating that more sophisticated regulation is embedded in the actual biosynthetic process. Subsequently, real-time PCR was used to analyze the transcription level of each *nor* gene in both *S. orinoci* and *S. lividans*. It was shown that most of the enzymes involved in spectinabilin biosynthesis were expressed at extremely low levels in *S. lividans* even in the absence of NorD repression. When referenced to the level of expression of *hrdB*, *norJ*, *norG*, and *norH* have more than 40-fold higher levels of transcription in *S. orinoci* than in *S. lividans* (Figure 2b). Such repressed transcription of multiple genes could explain the silencing of spectinabilin biosynthesis in the heterologous host.

Because understanding the regulation hierarchy remains an overwhelming challenge, the silent *nor* pathway in the heterologous host serves as a perfect test bed for our scaffold design in refactoring gene clusters. Therefore, nine strong constitutive promoters from our collection were used to drive the expression of the *nor* genes except for *norD* and *norG* (Figure 3a). NorG is the first enzyme in the pathway, converting chorismate from the shikimate pathway to *p*-aminobenzoic acid. For *norG*, the hyperinducible promoter nitA<sub>p</sub> induced by the cheap *ε*-caprolactam<sup>35</sup> was used. Although spectinabilin is not toxic to *S. lividans*, we included an inducible promoter to demonstrate a more generally applicable design because of the consideration that many natural products have biological activities and might be toxic to the heterologous host. From another point of view, in nature, secondary metabolites



**Figure 3.** (a) Refactored spectinabilin pathway. (b) Liquid chromatography–mass spectrometry analysis of the extracts from the *S. lividans* strain carrying the refactored *nor* pathway and the native *nor* pathway, respectively. The peak labeled with a star indicated the target product peak. (c) Real-time PCR analysis of transcription of the *nor* gene in the refactored pathway in *S. lividans* (for the purpose of direct comparison, the real-time PCR analysis of the *nor* gene transcription in *S. orinoci* illustrated in Figure 2b was incorporated).

are mostly synthesized in the stationary phase, when native producers do not need many resources to support their primary metabolism. Therefore, the competition between primary metabolite biosynthesis and the target secondary metabolite biosynthesis for the precursors is at least partially relieved by the inclusion of an inducible promoter. The preparation of the three helper modules was previously reported.<sup>12</sup> The refactored spectinabilin gene cluster was built by PCR-amplifying each *nor* gene (except for *norD*) and its downstream intergenetic sequence and plugging them into the scaffold (Figures S5 and S6 of the Supporting Information). As a result, the refactored pathway was successfully activated and produced spectinabilin in *S. lividans* (Figure 3b and Figure S7 of the Supporting Information), with a titer of  $105 \pm 21 \mu\text{g/L}$ . We further investigated the transcription of the *nor* genes in the refactored cluster and confirmed that all the genes driven by the strong promoters were turned on at high levels (Figure 3c), in contrast to the low levels of most genes in the intact cluster (Figure 2b). Among them, only *norG*, expression of which was driven by the inducible promoter nitA<sub>p</sub>, had a relatively low transcription level.

In contrast to the traditional strategies that require investigators to individually examine the transcriptional regulation specific to the cluster and the host, here we describe a generic design to replace the sophisticated regulatory elements with a set of characterized ones. The key steps are as follows.

(i) Determine an ideal expression host, which in general should be closely related to the native producer to provide both necessary precursors and a similar environment for protein translation and folding. The choice of the expression host should also be made on the basis of the availability of genetic

tools for manipulating the organism. For example, *S. lividans* was used as a host to express clusters from other *Streptomyces* species,<sup>16–20</sup> and *S. cerevisiae* was used to express clusters from fungi.<sup>36,37</sup>

(ii) Identify a set of strong constitutive promoters. Except for *E. coli* and *S. cerevisiae*, most organisms do not have such a set of promoters reported in the literature for ready usage. Because of the development of real-time PCR and RNA-seq, strong constitutive promoters can be rapidly identified from the upstream regions of the housekeeping genes in the selected expression host. The culturing condition for promoter identification will be subsequently used to produce the target compound such that the transcription of each pathway gene is forced to be on. As we demonstrated here, promoters from closely related organisms are very likely to be recognized by the transcription machinery in the selected host. These promoters could undergo unknown regulations in the target host and, thus, might not be completely constitutive, but at least transcription of the downstream genes should be consistently observed. Note that we first used real-time PCR analysis to identify two strong promoters from the heterologous host and then relied on the XylE activity assay to confirm the activities of all the candidate promoters. We did this mainly because of the following considerations. (1) The promoters we identified were from various sources; therefore, they need to be cloned upstream of a single reporter gene to have a fair comparison of their activities. (2) The time course of the mRNA level varies from promoter to promoter. As shown in Figure S8a of the Supporting Information, the mRNA levels of *gapdh* and *rpsL* genes in *S. griseus* were very high in the samples collected at 12 h, the mRNA level of *rpsL* decreased significantly afterward, and the mRNA level of *gapdh* continued to climb up until 24 h.

If the strengths were evaluated on the basis of the samples collected at the time points after 12 h, we probably would have missed  $rpsL_p(SG)$ , although its high transcription efficiency at an early stage already had led to sufficient protein expression (Figure 1b). For the same reason, the order of the promoter strength measured on the basis of the mRNA level will probably not be consistent with that measured on the basis of the protein assay (Figure 1b and Figure S8b of the Supporting Information). In addition, note that we attempt to provide a basic guideline to awaken a pathway without knowing its cotranscribed groups of genes. It is also reasonable to consider inserting one or two promoters in front of cotranscribed groups of genes to turn on the expression if such knowledge is available.

(iii) PCR-amplify each pathway gene and its downstream intergenic region from the target gene cluster using primers that will generate overlaps between adjacent fragments and then refactor the gene cluster using DNA assembler in *S. cerevisiae*. The resulting refactored gene cluster will maintain its native termination elements if they exist, but its transcription is fully controlled by the inserted heterologous promoters. Compared to promoter identification, it is more challenging to identify a set of terminators, and we also tested a few online bioinformatics tools but did not obtain reliable predictions. Certainly, an entire intergenic region can possibly contain another potential promoter or a potential terminator for the downstream gene depending on the direction of that gene. However, in the refactored gene cluster, a strong promoter is placed directly upstream of each pathway gene, which will control the transcription such that the question of whether another independent regulatory element exists upstream of this strong promoter becomes trivial.

In addition, several other DNA assembly methods such as Gibson cloning,<sup>38</sup> RecET-mediated direct cloning,<sup>39,40</sup> and reiterative recombination,<sup>41</sup> all based on *in vivo* or *in vitro* homologous recombination, were reported recently. On the basis of our experience with all these methods, DNA assembler shows higher assembly accuracy with an easy protocol for pathways of >20 kb, especially when the number of fragments to be assembled is >10. This is likely because *S. cerevisiae* has a full set of machinery responsible for high-fidelity homology recombination. Pathways up to 50 kb can be assembled routinely within 1–2 weeks with an efficiency of 30–100%.<sup>12</sup> Such an *in vivo* homologous recombination-based assembly has also been used to assemble molecules as large as a genome.<sup>42</sup> Combined with the rapid promoter identification mentioned above, the method we demonstrate here allows pathway design and manipulation in any desired organism as easy as in *E. coli* and *S. cerevisiae*. Codons can be optimized in this step if needed.

(iv) Express the cluster in the selected host and identify the product. The expression of the pathway genes can be confirmed via real-time PCR, sodium dodecyl sulfate–polyacrylamide gel electrophoresis, and Western blotting (if a gene is tagged). Metabolites extracted from the expression host carrying the refactored cluster are usually analyzed by liquid chromatography and compared with those extracted from the host lacking the exogenous pathway or carrying a nonfunctional pathway (e.g., a pathway with an essential gene deleted or mutated). Compounds appearing as new peaks are purified and subjected to mass spectrometry or NMR analysis for structure clarification.

This plug-and-play design sheds light on studying cryptic pathways for which the corresponding products have not been identified.<sup>43–45</sup> Over the past two decades, the complete genome sequences of more than 2000 organisms have been determined, with more than 10000 organisms being sequenced. Genome mining has revealed that the natural products that have been characterized are merely “the tip of the iceberg”, while plenty of metabolites await discovery. For example, genome mining of *Streptomyces*,<sup>46,47</sup> myxobacteria,<sup>48</sup> cyanobacteria,<sup>49,50</sup> and fungi<sup>51</sup> revealed the presence of many cryptic pathways involved in secondary metabolite production, while these strains were previously known to produce only a few compounds before their genomes were sequenced. The complex regulation embedded in natural product biosynthesis always hampers the discovery of novel natural products. The traditional methods for eliciting pathway expression suffer from the laborious process, lack general applicability, and repeatedly identify compounds that have already been identified. The strategy we present here offers an alternative for activating cryptic pathways identified through genome mining.

Like other existing approaches in the natural product research field, the plug-and-play scaffold does not solve all the problems and needs to be further improved in a few respects. For example, in some cases, the precursors are not abundant or even missing from the expression host, while in other cases, pathway expression could be far from the balancing point so the production titer is low. Incorporation of the genes encoding precursor synthesis into the refactored pathway could be a viable approach if those genes can be identified. It has been widely accepted that high-level transcription does not always guarantee a high titer. As shown in this study, the titer of spectinabilin in the native producer is actually much higher [ $>1$  mg/L (see Figure S7b of the Supporting Information)]. Note that the decision for making promoter and gene pairs was arbitrary, and we did not follow the strength order of the native promoters to pair the heterologous promoters with the *nor* genes. This was mainly because the dynamics of the native promoters was not similar to that of the heterologous promoters. Even for the same heterologous promoter, switching the downstream gene from *xylE* to a *nor* gene also resulted in very different mRNA levels partially because of the stability of mRNA (Figure 3c and Figure S8b of the Supporting Information). Moreover, even if we could match the strength order of the heterologous promoters with that of the native promoters at a certain time point, the order would not be maintained for other time points. That is the exact reason why we decided not to decipher the sophisticated regulation specific to each gene cluster. Instead, our strategy is to first use strong promoters to activate a target gene cluster and determine the chemical structure of the product and then rely on other pathway engineering strategies to balance the flux to improve the titer and yield of the product. If a high-throughput screening method is available, the expression of each pathway gene can be fine-tuned by setting up a promoter library with varying strengths to control gene transcription or alternatively creating a RBS library or an intergenic sequence library to post-transcriptionally control the expression, and the desired pathway variants carrying a balanced flux can be identified from the pool. Similar strategies for coordinating transcriptional or post-transcriptional processes have been successfully used in pathway optimization in *S. cerevisiae* and *E. coli*.<sup>52–54</sup> Now that a natural product gene cluster can be easily built *de novo* in many other expression hosts, a similar idea can be applied.

Moreover, modular design can be further applied to generate a series of distinct inducible promoters controlled by a single inducer, e.g., fusing the DNA binding sequence of the repressor, NitR (Figure S4 of the Supporting Information),<sup>35</sup> with other constitutive promoters used in refactoring the spectinabilin gene cluster, such that pathway genes are turned on at the same time. Lastly, our long-term goal is to build a high-throughput platform for natural product discovery. With the cost of chemical synthesis of DNA being continuously reduced, refactored gene clusters can be completely designed *in silico* and directly synthesized. Transformation of thousands of such gene clusters and screening the resulting library in a high-throughput manner should lead to the discovery of many interesting compounds. Regardless of subsequent improvements, our plug-and-play scaffold design is well-suited for *de novo* gene cluster assembly, rapid heterologous expression of biosynthetic gene clusters in tractable hosts, and mining the vast amount of genome sequence data for applications in secondary metabolite discovery.

## METHODS

**Materials and Reagents.** *S. lividans* 66 and *S. orinoci* were obtained from the Agricultural Research Service Culture Collection (Peoria, IL). Plasmid pAE4 and *E. coli* strain WM6026 were gifts from W. Metcalf (University of Illinois at Urbana-Champaign). Complete sequences of plasmids and details of strain and plasmid construction can be obtained by request from the authors. Plasmid pRS416 was purchased from New England Biolabs (Beverly, MA). Failsafe 2× premix buffer G was purchased from EPICENTRE Biotechnologies (Madison, WI), which was used as the reagent to amplify pathway fragments. Synthetic complete drop-out medium lacking uracil (SC-Ura) was used to select transformants containing the assembled pathways, and *S. cerevisiae* HZ848 (*MAT $\alpha$* , *ade2-1*,  $\Delta$ *ura3*, *his3-11, 15*, *trp1-1*, *leu2-3, 112* and *can1-100*) was used as the host for DNA assembly.

**Streptomyces Cultivation, RNA Extraction, and Real-Time PCR Analysis.** A seed culture was grown in MYG medium (4 g/L yeast extract, 10 g/L malt extract, and 4 g/L glucose) at 30 °C while being shaken constantly (250 rpm) until saturation was reached. The seed culture was inoculated into fresh MYG at a 1:100 ratio. For promoter screening and Xyle assay, cultures were collected at appropriate times and the total RNA was extracted using the RNeasy Mini Kit (Qiagen, Valencia, CA). Reverse transcription was conducted using the First-strand cDNA Synthesis Kit (Roche Applied Science, Indianapolis, IN). Real-time PCR was performed with SYBR Green PCR Master Mix on the 7900HT Fast Real-Time PCR System (Applied Biosystems, Carlsbad, CA). For investigation of *nor* gene expression in the refactored pathway,  $\epsilon$ -caprolactam was added to the cultures at a concentration of 1 g/L after 12 h, and samples were collected after an additional 24 h. For analysis of *nor* gene expression in the intact cluster in either *S. orinoci* or *S. lividans*, samples were collected after 36 h. The endogenous gene, *hrdB*, encoding RNA polymerase sigma factor, was used as the internal control for promoter screening. The expression of other candidate genes was normalized by the expression of the control. Data were analyzed with SDS2.4 (Applied Biosystems).

**Promoter Cloning.** For *gapdh<sub>p</sub>* (SG) and *rpsL<sub>p</sub>* (SG), the genomic DNA of *S. griseus* was isolated using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI) and the two target promoters were PCR-amplified from the isolated

genome. Promoters from other hosts were obtained through primer splicing, in which 6–10 overlapping oligonucleotides designed on the basis of the sequence of each promoter are joined through overlap extension PCR. The resulting promoters were further spliced with the amplified *xyleE* and cloned into the pAE4 vector, which is a *Streptomyces*–*E. coli* shuttle vector and also a *Streptomyces* integration vector (Figure S3 of the Supporting Information). The Xyle assay was performed according to the protocol described previously.<sup>20</sup>

**Pathway Refactoring and Yeast Transformation.** Pathway fragments were amplified from the genomic DNA of *S. orinoci*. The primer sequences are listed in Table S3 of the Supporting Information. The *S. cerevisiae* helper fragment was amplified from plasmid pRS416, whereas the *E. coli* helper fragment and the *S. lividans* helper fragment were amplified from pAE4. The PCR products were individually gel-purified from 0.7% agarose; 200–300 ng of each individual fragment was mixed and precipitated with ethanol. The resulting DNA pellet was air-dried and resuspended in 4  $\mu$ L of Milli-Q doubly deionized water. The previously reported two-step assembly strategy<sup>12</sup> was used to refactor the 42.6 kb spectinabilin gene cluster. To construct the three intermediate plasmids carrying the partial spectinabilin biosynthetic pathway (Figures S5 and S6 of the Supporting Information), the concentrated mixture of DNA was electroporated into *S. cerevisiae* using the protocol reported previously.<sup>13</sup> To construct the full-length spectinabilin pathway, the three intermediate plasmids were subjected to *AvrII* and *SspI* digestion and the released intermediate pathway fragments were combined with the master helper fragment (Figures S5 and S6 of the Supporting Information). After being concentrated, the mixture was transformed into *S. cerevisiae* using the lithium acetate/single-stranded carrier DNA/polyethylene glycol (PEG) method.<sup>55</sup>

**Restriction Digestion Analysis.** Colonies were randomly picked to SC-Ura liquid medium and grown for 1 day, after which the plasmids from yeast were isolated using the Zymoprep II Yeast Plasmid Miniprep kit (Zymo Research). Yeast plasmids were transformed into *E. coli* strain BW25141 and selected on Luria broth (LB) agar plates supplemented with 50  $\mu$ g/mL apramycin. Colonies were inoculated into 5 mL of LB medium supplemented with 50  $\mu$ g/mL apramycin, and plasmids were isolated from the liquid culture. Plasmids isolated from *E. coli* were then subjected to restriction digestion. Usually, one or two enzymes cutting the target molecule at multiple sites were chosen. The reaction mixtures were loaded onto 0.7% agarose gels to check for the correct restriction digestion pattern by DNA electrophoresis.

**Heterologous Expression in *S. lividans*.** The verified clones were transformed into *E. coli* WM6026<sup>16</sup> and selected on LB medium supplemented with 19  $\mu$ g/mL 2,6-diaminopimelic acid and 50  $\mu$ g/mL apramycin agar plates. These transformants were then used as the donors for conjugal transfer of the assembled plasmids to *S. lividans* 66 following the protocol described previously.<sup>16</sup> *S. lividans* exconjugants were picked and restreaked on ISP2 plates supplemented with 50  $\mu$ g/mL apramycin and allowed to grow for 2 days. A single colony was inoculated into 10 mL of MYG medium supplemented with 50  $\mu$ g/mL apramycin and grown at 30 °C for 2 days as a seed culture, of which 2.5 mL was subsequently inoculated into 250 mL of fresh MYG medium and grown for the appropriate times. For expression of the refactored pathway,  $\epsilon$ -caprolactam was added at a concentration of 1 g/L after 12 h, and samples were collected at appropriate times afterward.

**Liquid Chromatography–Mass Spectrometry Analysis.** Cultures were cleared of cells by centrifugation. The supernatants were extracted with an equal volume of ethyl acetate and concentrated 1000-fold before high-performance liquid chromatography (HPLC) analysis. HPLC was performed on an Agilent 1100 series LC/MSD XCT plus ion trap mass spectrometer (Agilent, Palo Alto, CA) with an Agilent SB-C18 reverse-phase column. HPLC parameters for detection of spectinabilin were as follows: solvent A, 1% acetic acid in water; solvent B, acetonitrile; gradient, 10% B for 5 min, to 100% B in 10 min, hold at 100% B for additional 5 min, return to 10% B over 10 min, and finally maintained at 10% B for 7 min; flow rate, 0.3 mL/min; detection by UV spectroscopy at 367 nm. Under such conditions, spectinabilin is eluted at 20.2 min. Mass spectra were acquired in ultra scan mode using electrospray ionization with positive polarity. The MS system was operated using a drying temperature of 350 °C, a nebulizer pressure of 35 psi, a drying gas flow of 8.5 L/min, and a capillary voltage of 4500 V.

## ■ ASSOCIATED CONTENT

### ● Supporting Information

This information is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Phone: 217 333-2631. Fax: 217 333-5052. E-mail: zhao5@illinois.edu.

### Present Address

<sup>||</sup>Z.S.: Department of Chemical and Biological Engineering, 4140 Biorenewables Research Laboratory, Iowa State University, Ames, IA 50010.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the National Academies Keck Futures Initiative on Synthetic Biology (NAKFI SB13).

## ■ REFERENCES

- (1) 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, 311–335.
- (2) Li, J. W., and Vederas, J. C. (2009) Drug discovery and natural products: End of an era or an endless frontier? *Science* 325, 161–165.
- (3) Bibb, M., and Hesketh, A. (2009) Analyzing the regulation of antibiotic production in streptomycetes. *Methods Enzymol.* 458, 93–116.
- (4) van Wezel, G. P., McKenzie, N. L., and Nodwell, J. R. (2009) Applying the genetics of secondary metabolism in model actinomycetes to the discovery of new antibiotics. *Methods Enzymol.* 458, 117–141.
- (5) Bibb, M. J. (2005) Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.* 8, 208–215.
- (6) Brakhage, A. A., Schuemann, J., Bergmann, S., Scherlach, K., Schroeckh, V., and Hertweck, C. (2008) Activation of fungal silent gene clusters: A new avenue to drug discovery. *Prog. Drug Res.* 66 (1), 3–12.
- (7) Bergmann, S., Schumann, J., Scherlach, K., Lange, C., Brakhage, A. A., and Hertweck, C. (2007) Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. *Nat. Chem. Biol.* 3, 213–217.
- (8) Wendt-Pienkowski, E., Huang, Y., Zhang, J., Li, B., Jiang, H., Kwon, H., Hutchinson, C. R., and Shen, B. (2005) Cloning, sequencing, analysis, and heterologous expression of the fredericamycin biosynthetic gene cluster from *Streptomyces griseus*. *J. Am. Chem. Soc.* 127, 16442–16452.
- (9) Baltz, R. H. (2010) *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. *J. Ind. Microbiol. Biotechnol.* 37, 759–772.
- (10) Komatsu, M., Uchiyama, T., Omura, S., Cane, D. E., and Ikeda, H. (2010) Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2646–2651.
- (11) Rodriguez, E., Hu, Z., Ou, S., Volchegursky, Y., Hutchinson, C. R., and McDaniel, R. (2003) Rapid engineering of polyketide overproduction by gene transfer to industrially optimized strains. *J. Ind. Microbiol. Biotechnol.* 30, 480–488.
- (12) Shao, Z., Luo, Y., and Zhao, H. (2011) Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. *Mol. Biosyst.* 7, 1056–1059.
- (13) Shao, Z., and Zhao, H. (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* 37, e16.
- (14) Newman, D. J., Cragg, G. M., and Snader, K. M. (2000) The influence of natural products upon drug discovery. *Nat. Prod. Rep.* 17, 215–234.
- (15) Hopwood, D. A. (2007) *Streptomyces in Nature and Medicine: The Antibiotic Makers*, Oxford University Press, Inc., New York.
- (16) Woodyer, R. D., Shao, Z., Thomas, P. M., Kelleher, N. L., Blodgett, J. A., Metcalf, W. W., van der Donk, W. A., and Zhao, H. (2006) Heterologous production of fosfomicin and identification of the minimal biosynthetic gene cluster. *Chem. Biol.* 13, 1171–1182.
- (17) Eliot, A. C., Griffin, B. M., Thomas, P. M., Johannes, T. W., Kelleher, N. L., Zhao, H., and Metcalf, W. W. (2008) Cloning, expression, and biochemical characterization of *Streptomyces rubellomurinus* genes required for biosynthesis of antimalarial compound FR900098. *Chem. Biol.* 15, 765–770.
- (18) Felnagle, E. A., Rondon, M. R., Berti, A. D., Crosby, H. A., and Thomas, M. G. (2007) Identification of the biosynthetic gene cluster and an additional gene for resistance to the antituberculosis drug capreomycin. *Appl. Environ. Microbiol.* 73, 4162–4170.
- (19) Feng, Z., Wang, L., Rajski, S. R., Xu, Z., Coeffet-LeGal, M. F., and Shen, B. (2009) Engineered production of iso-migrastatin in heterologous *Streptomyces* hosts. *Bioorg. Med. Chem.* 17, 2147–2153.
- (20) Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000) *Practical Streptomyces Genetics*, John Innes Foundation, Norwich, U.K.
- (21) Wagner, N., Osswald, C., Biener, R., and Schwartz, D. (2009) Comparative analysis of transcriptional activities of heterologous promoters in the rare actinomycete *Actinoplanes friuliensis*. *J. Biotechnol.* 142, 200–204.
- (22) Schmitt-John, T., and Engels, J. W. (1992) Promoter constructions for efficient secretion expression in *Streptomyces lividans*. *Appl. Microbiol. Biotechnol.* 36, 493–498.
- (23) Ahmad, I., Shim, W. Y., Jeon, W. Y., Yoon, B. H., and Kim, J. H. (2012) Enhancement of xylitol production in *Candida tropicalis* by co-expression of two genes involved in pentose phosphate pathway. *Bioprocess Biosyst. Eng.* 35, 199–204.
- (24) Hong, W. K., Kim, C. H., Heo, S. Y., Luo, L., Oh, B. R., and Seo, J. W. (2010) Enhanced production of ethanol from glycerol by engineered *Hansenula polymorpha* expressing pyruvate decarboxylase and aldehyde dehydrogenase genes from *Zymomonas mobilis*. *Biotechnol. Lett.* 32, 1077–1082.
- (25) Ahn, J., Hong, J., Lee, H., Park, M., Lee, E., Kim, C., Choi, E., Jung, J., and Lee, H. (2007) Translation elongation factor 1- $\alpha$  gene from *Pichia pastoris*: Molecular cloning, sequence, and use of its promoter. *Appl. Microbiol. Biotechnol.* 74, 601–608.
- (26) Sun, J., Shao, Z., Zhao, H., Nair, N., Wen, F., and Xu, J. H. (2012) Cloning and characterization of a panel of constitutive

promoters for applications in pathway engineering in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 109, 2082–2092.

(27) Krasny, L., Vacik, T., Fucik, V., and Jonak, J. (2000) Cloning and characterization of the *str* operon and elongation factor Tu expression in *Bacillus stearothermophilus*. *J. Bacteriol.* 182, 6114–6122.

(28) Jia, Y. L., Li, S. K., Allen, G., Feng, S. Y., and Xue, L. X. (2012) A Novel Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) Promoter for Expressing Transgenes in the Halotolerant Alga *Dunaliella salina*. *Curr. Microbiol.* 64, 506–513.

(29) Suarez, C. E., Norimine, J., Lacy, P., and McElwain, T. F. (2006) Characterization and gene expression of *Babesia bovis* elongation factor-1 $\alpha$ . *Int. J. Parasitol.* 36, 965–973.

(30) Traitcheva, N., Jenke-Kodama, H., He, J., Dittmann, E., and Hertweck, C. (2007) Non-colinear polyketide biosynthesis in the aureothin and neo-aureothin pathways: An evolutionary perspective. *ChemBioChem* 8, 1841–1849.

(31) Isaka, M., Jaturapat, A., Kramyu, J., Tanticharoen, M., and Thebtaranonth, Y. (2002) Potent in vitro antimalarial activity of metacycloprodigiosin isolated from *Streptomyces spectabilis* BCC 4785. *Antimicrob. Agents Chemother.* 46, 1112–1113.

(32) Kakinuma, K., Hanson, C. A., and Rinehart, K. L., Jr. (1976) Spectinabilin, a new nitro-containing metabolite isolated from *Streptomyces spectabilis*. *Tetrahedron* 32, 217–222.

(33) Choi, Y. S., Johannes, T. W., Simurdiak, M., Shao, Z., Lu, H., and Zhao, H. (2010) Cloning and heterologous expression of the spectinabilin biosynthetic gene cluster from *Streptomyces spectabilis*. *Mol. Biosyst.* 6, 336–338.

(34) Traitcheva, N. (2006) Molecular Analysis of the Neo-aureothin Biosynthesis Gene Cluster from *Streptomyces orinoci* HKI 0260: A Model System for the Evolution of Bacterial Polyketide Synthases. Ph.D. Thesis, Leibniz-Institute for Natural Product Research and Infection Biology (HKI), Jena, Germany.

(35) Herai, S., Hashimoto, Y., Higashibata, H., Maseda, H., Ikeda, H., Omura, S., and Kobayashi, M. (2004) Hyper-inducible expression system for streptomycetes. *Proc. Natl. Acad. Sci. U.S.A.* 101, 14031–14035.

(36) Wawrzyn, G. T., Bloch, S. E., and Schmidt-Dannert, C. (2012) Discovery and characterization of terpenoid biosynthetic pathways of fungi. *Methods Enzymol.* 515, 83–105.

(37) Ishiuchi, K., Nakazawa, T., Ookuma, T., Sugimoto, S., Sato, M., Tsunematsu, Y., Ishikawa, N., Noguchi, H., Hotta, K., Moriya, H., and Watanabe, K. (2012) Establishing a new methodology for genome mining and biosynthesis of polyketides and peptides through yeast molecular genetics. *ChemBioChem* 13, 846–854.

(38) Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345.

(39) Fu, J., Bian, X., Hu, S., Wang, H., Huang, F., Seibert, P. M., Plaza, A., Xia, L., Muller, R., Stewart, A. F., and Zhang, Y. (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* 30, 440–446.

(40) Cobb, R. E., and Zhao, H. M. (2012) Direct cloning of large genomic sequences. *Nat. Biotechnol.* 30, 405–406.

(41) Winkler, L. M., and Cornish, V. W. (2011) Reiterative recombination for the in vivo assembly of libraries of multigene pathways. *Proc. Natl. Acad. Sci. U.S.A.* 108, 15135–15140.

(42) Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., Stockwell, T. B., Brownley, A., Thomas, D. W., Algire, M. A., Merryman, C., Young, L., Noskov, V. N., Glass, J. I., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 319, 1215–1220.

(43) Gross, H. (2007) Strategies to unravel the function of orphan biosynthesis pathways: Recent examples and future prospects. *Appl. Microbiol. Biotechnol.* 75, 267–277.

(44) Challis, G. L. (2008) Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology* 154, 1555–1569.

(45) Zerikly, M., and Challis, G. L. (2009) Strategies for the Discovery of New Natural Products by Genome Mining. *ChemBioChem* 10, 625–633.

(46) Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., Yamashita, A., Hattori, M., and Horinouchi, S. (2008) Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J. Bacteriol.* 190, 4050–4060.

(47) Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M. (2001) Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12215–12220.

(48) Bode, H. B., and Muller, R. (2006) Analysis of myxobacterial secondary metabolism goes molecular. *J. Ind. Microbiol. Biotechnol.* 33, 577–588.

(49) Abed, R. M., Dobretsov, S., and Sudesh, K. (2009) Applications of cyanobacteria in biotechnology. *J. Appl. Microbiol.* 106, 1–12.

(50) Singh, S., Kate, B. N., and Banerjee, U. C. (2005) Bioactive compounds from cyanobacteria and microalgae: An overview. *Crit. Rev. Biotechnol.* 25, 73–95.

(51) Brakhage, A. A., and Schroeckh, V. (2011) Fungal secondary metabolites: Strategies to activate silent gene clusters. *Fungal Genet. Biol.* 48, 15–22.

(52) Du, J., Yuan, Y., Si, T., Lian, J., and Zhao, H. (2012) Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Res.* 40, e142.

(53) Pflieger, B. F., Pitera, D. J., Smolke, C. D., and Keasling, J. D. (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.* 24, 1027–1032.

(54) Salis, H. M. (2011) The ribosome binding site calculator. *Methods Enzymol.* 498, 19–42.

(55) Gietz, R. D., and Woods, R. A. (2006) Yeast transformation by the LiAc/SS Carrier DNA/PEG method. *Methods Mol. Biol.* 313, 107–120.

(56) Kim, S. H., Lee, H. N., Kim, H. J., and Kim, E. S. (2011) Transcriptome analysis of an antibiotic downregulator mutant and synergistic Actinorhodin stimulation via disruption of a precursor flux regulator in *Streptomyces coelicolor*. *Appl. Environ. Microbiol.* 77, 1872–1877.

(57) Gomez, C., Horna, D. H., Olano, C., Palomino-Schatzlein, M., Pineda-Lucena, A., Carbajo, R. J., Brana, A. F., Mendez, C., and Salas, J. A. (2011) Amino acid precursor supply in the biosynthesis of the RNA polymerase inhibitor streptolydigin by *Streptomyces lydicus*. *J. Bacteriol.* 193, 4214–4223.

(58) Nieselt, K., Battke, F., Herbig, A., Bruheim, P., Wentzel, A., Jakobsen, O. M., Sletta, H., Alam, M. T., Merlo, M. E., Moore, J., Omara, W. A., Morrissey, E. R., Juarez-Hermosillo, M. A., Rodriguez-Garcia, A., Nentwich, M., Thomas, L., Iqbal, M., Legaie, R., Gaze, W. H., Challis, G. L., Jansen, R. C., Dijkhuizen, L., Rand, D. A., Wild, D. L., Bonin, M., Reuther, J., Wohlleben, W., Smith, M. C., Burroughs, N. J., Martin, J. F., Hodgson, D. A., Takano, E., Breitling, R., Ellingsen, T. E., and Wellington, E. M. (2010) The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. *BMC Genomics* 11, 10.

(59) Wilkinson, C. J., Hughes-Thomas, Z. A., Martin, C. J., Bohm, I., Mironenko, T., Deacon, M., Wheatcroft, M., Wirtz, G., Staunton, J., and Leadlay, P. F. (2002) Increasing the efficiency of heterologous promoters in actinomycetes. *J. Mol. Microbiol. Biotechnol.* 4, 417–426.