Published in final edited form as:
ACS Synth Biol. 2013 November 15; 2(11): 662–669. doi:10.1021/sb400058n.

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

Zengyi Shao[1], **Guodong Rao**[2], **Chun Li**[1], **Zhanar Abil**[3], **Yunzi Luo**[1], and **Huimin Zhao**[1,2,3,*]

[1]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801

[2]Departments of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801

[3]Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801

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 individual pathway is tightly controlled through unique regulation mechanism and hence often remains silent in the routine culturing conditions. To overcome the drawback 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

Introduction

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 (1, 2). The rapidly increasing number of sequenced genomes and metagenomes provide a tremendously rich source for discovery of gene clusters involved in biosynthesis of new compounds. However,

*To whom correspondence should be addressed. Phone: (217) 333-2631. Fax: (217) 333-5052. zhao5@illinois.edu.
Dr. Zengyi Shao: Assistant professor, Department of Chemical and Biological Engineering, 4140 Biorenewables Research Laboratory, Iowa State University, Ames, IA, 50010. All the other authors: 600 South Mathews Avenue, RAL221, University of Illinois at Urbana-Champaign, Urbana, IL 61801.

Supporting Information Available
This information is available free of charge via the Internet at http://pubs.acs.org/.

the discovery and economical production of natural products are hampered by our limited knowledge in manipulating most organisms, determining suitable conditions to elicit pathway expression, and producing sufficient compounds for structure identification.

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 (3–6). They interact with each other to form an extremely complex network in response to a variety of physiological and environmental signals. Existing approaches to study natural product biosynthetic clusters mainly include: (i) manipulating cell culture parameters, such as medium composition, to ensure expression of pathway-specific activator(s), presence of physiological and environmental co-inducers, or derepression of the genes repressed by repressor(s) (4, 6); (ii) engineering the regulation by expressing the pathway-specific regulator under a well-characterized promoter (7, 8); (iii) testing a variety of heterologous hosts to express the target cluster (9); (iv) silencing major secondary metabolite biosynthetic pathways in order to simplify product identification and relieve competition for key precursors (10); (v) utilizing industrial strains which have been set up for high-level production of specific compounds (11). All these strategies can only be applied on a case-by-case basis. Each gene cluster has its own unique regulatory mechanism that has to be examined individually in order 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 to decouple pathway expression from native sophisticated regulation cascades. Based on 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 are confirmed in the target expression host under the selected culture condition 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, 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 plus their downstream intergenic sequences in the target cluster can be individually amplified from the isolated genomic DNA if the native host is cultivable 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 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 (12, 13). Using such a plug-and-play scaffold, the sophisticated regulation embedded in individual clusters is removed and replaced by a set of regulation 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 above-mentioned plug-and-play strategy 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 (14, 15). *Streptomyces lividans*, a laboratory strain that has been used extensively for studying gene clusters from Streptomycetes was chosen as a target heterologous host owing to its high conjugative DNA transfer efficiency and the availability of several genetic manipulation tools (9, 16–20).

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 (Supporting Information Table S1). Using real-time PCR, two genes encoding glyceraldehyde-3-phosphate dehydrogenase (*gapdh*) and 30S ribosomal protein S12 (*rpsL*), stood out with much higher transcription levels than all the other genes in our fixed culturing condition (Figure 1a), indicating that their corresponding promoters could be very strong. The gapdh promoter, named as $gapdh_p$ (SG), is located upstream of the gapdh operon consisting of Gapdh, phosphoglycerate kinase (*pgk*) and triosephosphate isomerase (*tpiA*), the enzymes catalyzing the 6^{th}, 7^{th} and 5^{th} steps, respectively, in the glycolysis pathway; and the rpsL promoter, named as $rpsL_p$ (SG), resides upstream of another operon consisting of 30s ribosomal proteins S12 and S7, and elongation factor G and Tu (Supporting Information Figure S1). To ensure that these promoters can be used to drive the transcription of heterologous genes and also compare their activities with the ones reported in literature, the intergenic region between the gene located upstream and the gapdh operon (or the rpsL operon) were subsequently fused with the Streptomycete reporter gene, catechol 2, 3-dioxygenase (*xylE*), which catalyzes the conversion of colorless catechol to yellow-color 2-hydroxymuconic semialdehyde. The xylE activity assay (20) confirmed their much stronger activities than the control promoter, $ermE^*_p$, (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 (21, 22). The strong activities of $gapdh_p$ and the promoters of various translation elongation factors have also been observed in many other microorganisms, such as fungi, bacteria, micro-algae, and protozoa (23–29). Encouraged by

the strong activities of $gapdh_p$ and $rpsL_p$, we decided to examine their equivalents from other species. The promoters from different Streptomyces species share extremely high homologies with each other (Supporting Information Figure S2a), which is undesired because such high sequence similarities would cause severe deletions during DNA assembly in *S. cerevisiae*. Since Streptomycetes belong to the actinobacteria family, the promoters from other genera of actinobacteria could be possibly recognized by the transcription machinery in Streptomycetes as well. The gapdh and the rpsL operon were found to be highly conserved in the actinobacteria family, but the corresponding promoters are highly diversified (Supporting Information Figure S2b). Next, the potential $gapdh_p$ and $rpsL_p$ from 18 distinct actinobacteria (Supporting Information Figure S3) were cloned upstream of *xylE* for more quantitative comparison. As a result, 13 out 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 $ermE^*_p$ (Figure 1b and Supporting Information Table S2).

As proof of concept, the spectinabilin gene cluster from *Streptomyces orinoci* (30) was chosen as a model pathway. Spectinabilin, isolated from *Streptomyces spectabilis* and *Streptomyces orinoci*, is a nitro-phenyl substituted polyketide exhibiting antimalarial and antiviral activities (31, 32). Interestingly, although the catalytic proteins from the two spectinabilin clusters share very high homologies with each other, these two clusters undergo different regulations. The cluster from *S. spectabilis* (named as *spn* cluster) contains SpnD as an activator and the cluster from *S. orinoci* (named as *nor* cluster) contains NorD as a repressor (Figure 2a and Supporting Information Figure S4) (33). When the two clusters were cloned into *S. lividans*, only the *spn* cluster could heterologously produce spectinabilin under the laboratory conditions (33, 34). We first reconstructed the *nor* cluster excluding *norD* in *S. lividans*. However, no production of spectinabilin was observed either (data not shown), 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 expression of *hrdB*, *norJ*, *norG*, and *norH* have more than 40-fold higher 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 hyper-inducible promoter $nitA_p$ induced by the cheap ε-caprolactam (35) was used. Although spectinabilin is not toxic to *S. lividans*, we included an inducible promoter to demonstrate a more generally applicable design due to 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 are mostly synthesized in the stationary phase, when native producers do not

need a lot of 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 (12). The refactored spectinabilin gene cluster was built by PCR-amplifying each *nor* gene (except for *norD*) plus its downstream intergenetic sequence and plugging them into the scaffold (Supporting Information Figure S5 and S6). As a result, the refactored pathway was successfully activated and produced spectinabilin in *S. lividans* (Figure 3b and Supporting Information Figure S7), with a titer of 105±21 μg/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*, whose express was driven by the inducible promoter $nitA_p$, had a relative 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 described a generic design to replace the sophisticated regulatory elements with a set of characterized ones. The key steps include:

(i) Determine an ideal expression host, which in general should be closely related to the native producer in order to provide both necessary precursors and a similar environment for protein translation and folding. The choice of the expression host should also be made based on the availability of genetic tools to manipulate the organism. For example, *S. lividans* was used as a host to express clusters from other Streptomyces species (16–20) and *S. cerevisiae* was used to express clusters from fungi (36, 37).

(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 literature for ready usage. Thanks to 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 have a very high chance to be recognized by the transcription machinery in the selected host. These promoters could undergo unknown regulations in the target host, 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 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 in order to have a fair comparison of their activities. (2) The time course of the mRNA level varies from promoter to promoter. As shown in Supporting

Information Figure S8a, the mRNA levels of *gapdh* and *rpsL* genes in *S. griseus* were very high in the samples collected at 12 hours, but the mRNA level of *rpsL* decreased significantly afterwards while the mRNA level of *gapdh* continued to climb up until 24 hours. If the strengths were evaluated based on the samples collected at the time points after 12 hours, we probably would miss $rpsL_p$(SG) although its high transcription efficiency at an early stage already led to sufficient protein expression (Figure 1b). For the same reason, the order of the promoter strength measured based on mRNA level will probably not be consistent with that based on protein assay (Figure 1b and Supporting Information Figure S8b). In addition, note that we attempt to provide a basic guideline to awaken a pathway without knowing its co-transcribed groups of genes. It is also reasonable to consider inserting one or two promoters, in front of co-transcribed groups of genes to turn on the expression if such knowledge is available.

(iii) PCR-amplify each pathway gene plus 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 on-line 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 whether another independent regulatory element existing upstream of this strong promoter becomes trivial.

In addition, several other DNA assembly methods such as Gibson cloning (38), RecET mediated direct cloning (39, 40), and reiterative recombination (41), all based on *in vivo* or *in vitro* homologous recombination, were reported recently. Based on our experience with all these methods, DNA assembler shows higher assembly accuracy with an easy protocol for pathways larger than 20 kb, especially when the number of the fragments to be assembled is more than 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 week with an efficiency of 30–100% (12). Such an *in vivo* homologous recombination-based assembly has also been used to assemble molecules as large as a genome (42). Combined with the rapid promoter identification mentioned above, the method we propose here enables 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, SDS-PAGE and

Western blot (if a gene is tagged). Metabolites extracted from the expression host carrying the refactored cluster are usually analyzed by LC and compared with those extracted from the host lacking the exogenous pathway or carrying a non-functional 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 (43–45). Over the last two decades, the complete genome sequences of more than 2000 organisms have been determined, with more than 10,000 organisms being sequenced. Genome mining has revealed that the natural products which have been characterized are merely “the tip of the iceberg”, whilst plenty of metabolites await discovery. For example, genome mining of Streptomyces (46, 47), myxobacteria (48), cyanobacteria (49, 50) and fungi (51) 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 to elicit pathway expression suffer from the laborious process, lack general applicability, and repeatedly identify compounds that are already known. 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 need to be further improved in a few aspects. 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 for 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 Supporting Information Figure S7b). 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 due to the stability of mRNA (Figure 3c and Supporting Information Figure S8b). 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 to coordinate transcriptional or post-transcriptional processes have been successfully used in pathway optimization in *S. cerevisiae* and *E. coli* (52–54). 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 (Supporting Information Figure S4) (35) 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 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). Plasmids pAE4 and *E. coli* strain WM6026 were a gift from William Metcalf (University of Illinois, Urbana). Complete sequences of plasmids and details of strain and plasmid construction can be obtained by request from the authors. The plasmid pRS416 was purchased from New England Biolabs (Beverly, MA). Failsafe™ 2x 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*α, *ade*2-1, Δ*ura*3, *his*3-11, 15, *trp*1-1, *leu*2-3, 112 and *can*1-100) was used as the host for DNA assembly.

Streptomycetes 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 with constant shaking (250 rpm) until saturation, and inoculated into fresh MYG with a ratio of 1:100. 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 carried out 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, ε-caprolactam was added to the cultures at a concentration of 1 g/L after 12 hrs, and samples were collected after additional 24 hrs. For analysis of *nor* gene expression in the intact cluster in either *S. orinoci* or *S. lividans*, samples were collected

after 36 hrs. 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 was analyzed by the software SDS2.4 (Applied Biosystems).

Promoter cloning: For $gapdh_p$ (SG) and $rpsL_p$ (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 based on the sequence of each promoter are jointed through overlap extension PCR. The resulting promoters were further spliced with the amplified *xylE*, and cloned into pAE4 vector, which is a *Streptomyces*-*E. coli* shuttle vector and also a *Streptomyces* integration vector (Supporting Information Figure S3). XylE assay was performed according to the protocol described elsewhere (20).

Pathway refactoring and yeast transformation

Pathway fragments were amplified from the genomic DNA of *S. orinoci*. The primer sequences are listed in Supporting Information Table S3. The *S. cerevisiae* helper fragment was amplified from the 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 μL Milli-Q double deionized water. The previously reported two-step assembly strategy (12) was used to refactor the 42.6 kb spectinabilin gene cluster. To construct the three intermediate plasmids carrying the partial spectinabilin biosynthetic pathway (Supporting information Figure S5 and S6), the concentrated mixture of DNA was electroporated into *S. cerevisiae* using the protocol reported previously (13). To construct the full-length spectinabilin pathway, the three intermediate plasmids were subjected to *Avr*II and *Ssp*I digestion and the released intermediate pathway fragments were combined with the master helper fragment (Supporting information Figure S5 and S6). After being concentrated, the mixture was transformed to *S. cerevisiae* using the lithium acetate/single stranded carrier DNA/polyethylene glycol (PEG) method (55).

Restriction digestion analysis

Colonies were randomly picked to SC-Ura liquid media and grown for 1 day, after which the plasmids from yeast were isolated using Zymoprep II Yeast Plasmid Miniprep kit (Zymo Research, CA). Yeast plasmids were transformed to *E. coli* strain BW25141 and selected on Luria Broth (LB) agar plates supplemented with 50 μg/mL apramycin. Colonies were inoculated into 5 mL of LB media supplemented with 50 μg/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 to 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 to *E. coli* WM6026 (16) and selected on LB supplemented with 19 μg/mL 2,6-diaminopimelic acid and 50 μg/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 (16). *S. lividans* exconjugants were picked and restreaked on ISP2 plates supplemented with 50 μg/mL apramycin and allowed to grow for 2 days. A single colony was inoculated into 10 mL MYG medium supplemented with 50 μg/mL apramycin and grown at 30 °C for 2 days as a seed culture, of which 2.5 mL was subsequently inoculated to 250 mL fresh MYG medium and grown for appropriate times. For expressing the refactored pathway, ε-caprolactam was added at a concentration of 1 g/L after 12 hrs, and samples were collected at appropriate times afterward.

LC-MS 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, maintain at 100% B for 5 min, return to 10% B in 10 min and finally maintain 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 (ESI) 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^{-1}, and a capillary voltage of 4500 V.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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

References

1. Newman DJ, Cragg GM. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. Journal of Natural Products. 2012; 75:311–335. [PubMed: 22316239]
2. Li JW, Vederas JC. Drug discovery and natural products: end of an era or an endless frontier? Science. 2009; 325:161–165. [PubMed: 19589993]
3. Bibb M, Hesketh A. Chapter 4. Analyzing the regulation of antibiotic production in streptomycetes. Methods Enzymol. 2009; 458:93–116. [PubMed: 19374980]
4. van Wezel GP, McKenzie NL, Nodwell JR. Chapter 5. Applying the genetics of secondary metabolism in model actinomycetes to the discovery of new antibiotics. Methods Enzymol. 2009; 458:117–141. [PubMed: 19374981]
5. Bibb MJ. Regulation of secondary metabolism in streptomycetes. Current Opinion in Microbiology. 2005; 8:208–215. [PubMed: 15802254]

6. Brakhage AA, Schuemann J, Bergmann S, Scherlach K, Schroeckh V, Hertweck C. Activation of fungal silent gene clusters: a new avenue to drug discovery. Prog Drug Res. 2008; 66:1, 3–12. [PubMed: 18416304]
7. Bergmann S, Schumann J, Scherlach K, Lange C, Brakhage AA, Hertweck C. Genomics-driven discovery of PKS-NRPS hybrid metabolites from Aspergillus nidulans. Nat Chem Biol. 2007; 3:213–217. [PubMed: 17369821]
8. Wendt-Pienkowski E, Huang Y, Zhang J, Li B, Jiang H, Kwon H, Hutchinson CR, Shen B. Cloning, sequencing, analysis, and heterologous expression of the fredericamycin biosynthetic gene cluster from Streptomyces griseus. J Am Chem Soc. 2005; 127:16442–16452. [PubMed: 16305230]
9. Baltz RH. Streptomyces and Saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. J Ind Microbiol Biotechnol. 2010; 37:759–772. [PubMed: 20467781]
10. Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H. Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. Proc Natl Acad Sci U S A. 2010; 107:2646–2651. [PubMed: 20133795]
11. Rodriguez E, Hu Z, Ou S, Volchegursky Y, Hutchinson CR, McDaniel R. Rapid engineering of polyketide overproduction by gene transfer to industrially optimized strains. J Ind Microbiol Biotechnol. 2003; 30:480–488. [PubMed: 12698320]
12. Shao Z, Luo Y, Zhao H. Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. Mol Biosyst. 2011; 7:1056–1059. [PubMed: 21327279]
13. Shao Z, Zhao H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. 2009; 37:e16. [PubMed: 19074487]
14. Newman DJ, Cragg GM, Snader KM. The influence of natural products upon drug discovery. Natural Product Reports. 2000; 17:215–234. [PubMed: 10888010]
15. Hopwood, DA. Streptomyces in Nature and Medicine: The Antibiotic Makers. Oxford University Press, Inc.; New York, United States: 2007.
16. Woodyer RD, Shao Z, Thomas PM, Kelleher NL, Blodgett JA, Metcalf WW, van der Donk WA, Zhao H. Heterologous production of fosfomycin and identification of the minimal biosynthetic gene cluster. Chem Biol. 2006; 13:1171–1182. [PubMed: 17113999]
17. Eliot AC, Griffin BM, Thomas PM, Johannes TW, Kelleher NL, Zhao H, Metcalf WW. Cloning, expression, and biochemical characterization of Streptomyces rubellomurinus genes required for biosynthesis of antimalarial compound FR900098. Chem Biol. 2008; 15:765–770. [PubMed: 18721747]
18. Felnagle EA, Rondon MR, Berti AD, Crosby HA, Thomas MG. Identification of the biosynthetic gene cluster and an additional gene for resistance to the antituberculosis drug capreomycin. Applied and Environmental Microbiology. 2007; 73:4162–4170. [PubMed: 17496129]
19. Feng Z, Wang L, Rajski SR, Xu Z, Coeffet-LeGal MF, Shen B. Engineered production of iso-migrastatin in heterologous Streptomyces hosts. Bioorg Med Chem. 2009; 17:2147–2153. [PubMed: 19010685]
20. Tobias Kieser, MJB.; Buttner, Mark J.; Chater, Keith F.; Hopwood, DA. Practical Streptomyces Genetics. John Innes Foundation; Norwich, United Kingdom: 2000.
21. Wagner N, Osswald C, Biener R, Schwartz D. Comparative analysis of transcriptional activities of heterologous promoters in the rare actinomycete Actinoplanes friuliensis. Journal of Biotechnology. 2009; 142:200–204. [PubMed: 19464328]
22. Schmitt-John T, Engels JW. Promoter constructions for efficient secretion expression in Streptomyces lividans. Appl Microbiol Biotechnol. 1992; 36:493–498. [PubMed: 1368202]
23. Ahmad I, Shim WY, Jeon WY, Yoon BH, Kim JH. Enhancement of xylitol production in Candida tropicalis by co-expression of two genes involved in pentose phosphate pathway. Bioprocess and Biosystems Engineering. 2012; 35:199–204. [PubMed: 21969058]
24. Hong WK, Kim CH, Heo SY, Luo L, Oh BR, Seo JW. Enhanced production of ethanol from glycerol by engineered Hansenula polymorpha expressing pyruvate decarboxylase and aldehyde dehydrogenase genes from Zymomonas mobilis. Biotechnology Letters. 2010; 32:1077–1082. [PubMed: 20354759]

25. Ahn J, Hong J, Lee H, Park M, Lee E, Kim C, Choi E, Jung J, Lee H. Translation elongation factor 1-alpha gene from Pichia pastoris: molecular cloning, sequence, and use of its promoter. Applied Microbiology and Biotechnology. 2007; 74:601–608. [PubMed: 17124582]
26. Sun J, Shao Z, Zhao H, Nair N, Wen F, Xu JH. Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in Saccharomyces cerevisiae. Biotechnol Bioeng. 2012; 109:2082–2092. [PubMed: 22383307]
27. Krasny L, Vacik T, Fucik V, Jonak J. Cloning and characterization of the str operon and elongation factor Tu expression in Bacillus stearothermophilus. Journal of Bacteriology. 2000; 182:6114–6122. [PubMed: 11029432]
28. Jia YL, Li SK, Allen G, Feng SY, Xue LX. A Novel Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) Promoter for Expressing Transgenes in the Halotolerant Alga Dunaliella salina. Current Microbiology. 2012; 64:506–513. [PubMed: 22371187]
29. Suarez CE, Norimine J, Lacy P, McElwain TF. Characterization and gene expression of Babesia bovis elongation factor-1alpha. Int J Parasitol. 2006; 36:965–973. [PubMed: 16677650]
30. Traitcheva N, Jenke-Kodama H, He J, Dittmann E, Hertweck C. Non-colinear polyketide biosynthesis in the aureothin and neoaureothin pathways: an evolutionary perspective. Chembiochem. 2007; 8:1841–1849. [PubMed: 17763486]
31. Isaka M, Jaturapat A, Kramyu J, Tanticharoen M, Thebtaranonth Y. Potent in vitro antimalarial activity of metacycloprodigiosin isolated from Streptomyces spectabilis BCC 4785. Antimicrob Agents Chemother. 2002; 46:1112–1113. [PubMed: 11897600]
32. Kakinuma K, Hanson CA, RinehartJr KL. Spectinabilin, a new nitro-containing metabolite isolated from streptomyces spectabilis. Tetrahedron. 1976; 32:217–222.
33. Choi YS, Johannes TW, Simurdiak M, Shao Z, Lu H, Zhao H. Cloning and heterologous expression of the spectinabilin biosynthetic gene cluster from Streptomyces spectabilis. Mol Biosyst. 2010; 6:336–338. [PubMed: 20094652]
34. Traitcheva, N, PhD thesis. Department of biomolecular chemistry. Leibniz-Institute for Natural Product Research and Infection Biology (HKI); Jena. Germany: 2006. Molecular Analysis of the Neoaureothin Biosynthesis Gene Cluster from Streptomyces orinoci HKI 0260: a Model System for the Evolution of Bacterial Polyketide Synthases; p. 98
35. Herai S, Hashimoto Y, Higashibata H, Maseda H, Ikeda H, Omura S, Kobayashi M. Hyper-inducible expression system for streptomycetes. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101:14031–14035. [PubMed: 15377796]
36. Wawrzyn GT, Bloch SE, Schmidt-Dannert C. Discovery and characterization of terpenoid biosynthetic pathways of fungi. Methods Enzymol. 2012; 515:83–105. [PubMed: 22999171]
37. Ishiuchi K, Nakazawa T, Ookuma T, Sugimoto S, Sato M, Tsunematsu Y, Ishikawa N, Noguchi H, Hotta K, Moriya H, Watanabe K. Establishing a new methodology for genome mining and biosynthesis of polyketides and peptides through yeast molecular genetics. Chembiochem. 2012; 13:846–854. [PubMed: 22447505]
38. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods. 2009; 6:343–U341. [PubMed: 19363495]
39. Fu J, Bian X, Hu S, Wang H, Huang F, Seibert PM, Plaza A, Xia L, Muller R, Stewart AF, Zhang Y. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. Nat Biotechnol. 2012; 30:440–446. [PubMed: 22544021]
40. Cobb RE, Zhao HM. Direct cloning of large genomic sequences. Nature Biotechnology. 2012; 30:405–406.
41. Wingler LM, Cornish VW. Reiterative Recombination for the in vivo assembly of libraries of multigene pathways. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:15135–15140. [PubMed: 21876185]
42. Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison CA, Smith HO. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science. 2008; 319:1215–1220. [PubMed: 18218864]

43. Gross H. Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. Applied Microbiology and Biotechnology. 2007; 75:267–277. [PubMed: 17340107]
44. Challis GL. Mining microbial genomes for new natural products and biosynthetic pathways. Microbiology-Sgm. 2008; 154:1555–1569.
45. Zerikly M, Challis GL. Strategies for the Discovery of New Natural Products by Genome Mining. Chembiochem. 2009; 10:625–633. [PubMed: 19165837]
46. Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, Yamashita A, Hattori M, Horinouchi S. Genome sequence of the streptomycin-producing microorganism Streptomyces griseus IFO 13350. J Bacteriol. 2008; 190:4050–4060. [PubMed: 18375553]
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, Hattori M. Genome sequence of an industrial microorganism Streptomyces avermitilis: Deducing the ability of producing secondary metabolites. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98:12215–12220. [PubMed: 11572948]
48. Bode HB, Muller R. Analysis of myxobacterial secondary metabolism goes molecular. J Ind Microbiol Biotechnol. 2006; 33:577–588. [PubMed: 16491362]
49. Abed RM, Dobretsov S, Sudesh K. Applications of cyanobacteria in biotechnology. J Appl Microbiol. 2009; 106:1–12. [PubMed: 19191979]
50. Singh S, Kate BN, Banerjee UC. Bioactive compounds from cyanobacteria and microalgae: an overview. Crit Rev Biotechnol. 2005; 25:73–95. [PubMed: 16294828]
51. Brakhage AA, Schroeckh V. Fungal secondary metabolites – strategies to activate silent gene clusters. Fungal Genet Biol. 2011; 48:15–22. [PubMed: 20433937]
52. Du J, Yuan Y, Si T, Lian J, Zhao H. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. Nucleic Acids Res. 2012; 40:e142. [PubMed: 22718979]
53. Pfleger BF, Pitera DJ, Smolke CD, Keasling JD. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol. 2006; 24:1027–1032. [PubMed: 16845378]
54. Salis HM. The ribosome binding site calculator. Methods Enzymol. 2011; 498:19–42. [PubMed: 21601672]
55. Gietz RD, Woods RA. Yeast transformation by the LiAc/SS Carrier DNA/PEG method. Methods Mol Biol. 2006; 313:107–120. [PubMed: 16118429]
56. Kim SH, Lee HN, Kim HJ, Kim ES. Transcriptome analysis of an antibiotic downregulator mutant and synergistic Actinorhodin stimulation via disruption of a precursor flux regulator in Streptomyces coelicolor. Appl Environ Microbiol. 2011; 77:1872–1877. [PubMed: 21216912]
57. Gomez C, Horna DH, Olano C, Palomino-Schatzlein M, Pineda-Lucena A, Carbajo RJ, Brana AF, Mendez C, Salas JA. Amino acid precursor supply in the biosynthesis of the RNA polymerase inhibitor streptolydigin by Streptomyces lydicus. J Bacteriol. 2011; 193:4214–4223. [PubMed: 21665968]
58. Nieselt K, Battke F, Herbig A, Bruheim P, Wentzel A, Jakobsen OM, Sletta H, Alam MT, Merlo ME, Moore J, Omara WA, Morrissey ER, Juarez-Hermosillo MA, Rodriguez-Garcia A, Nentwich M, Thomas L, Iqbal M, Legaie R, Gaze WH, Challis GL, Jansen RC, Dijkhuizen L, Rand DA, Wild DL, Bonin M, Reuther J, Wohlleben W, Smith MC, Burroughs NJ, Martin JF, Hodgson DA, Takano E, Breitling R, Ellingsen TE, Wellington EM. The dynamic architecture of the metabolic switch in Streptomyces coelicolor. BMC Genomics. 2010; 11:10. [PubMed: 20053288]
59. Wilkinson CJ, Hughes-Thomas ZA, Martin CJ, Bohm I, Mironenko T, Deacon M, Wheatcroft M, Wirtz G, Staunton J, Leadlay PF. Increasing the efficiency of heterologous promoters in actinomycetes. Journal of Molecular Microbiology and Biotechnology. 2002; 4:417–426. [PubMed: 12125822]

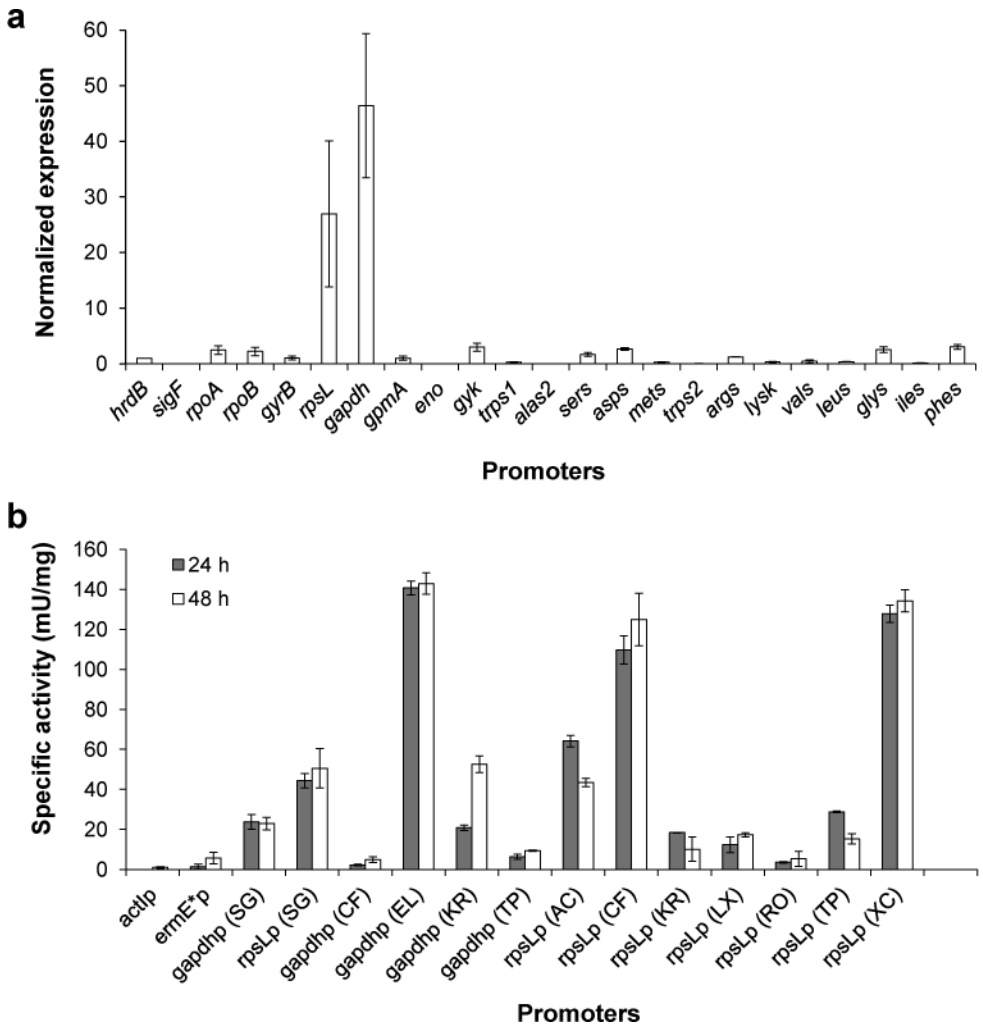


Figure 1.

Promoter screening for pathway refactoring in Streptomycetes. 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 (56–58), which was set to 1. The expressions of *gapdh* and *rpsL* were higher than the other genes at all the sampling points, with the 12-hour 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 $actI_p$ and $ermE^*_p$ are the two commonly used promoters reported in literature (21, 22, 59). The two letters in parentheses represent the names of individual actinomycetes (Supporting Information Table S2).

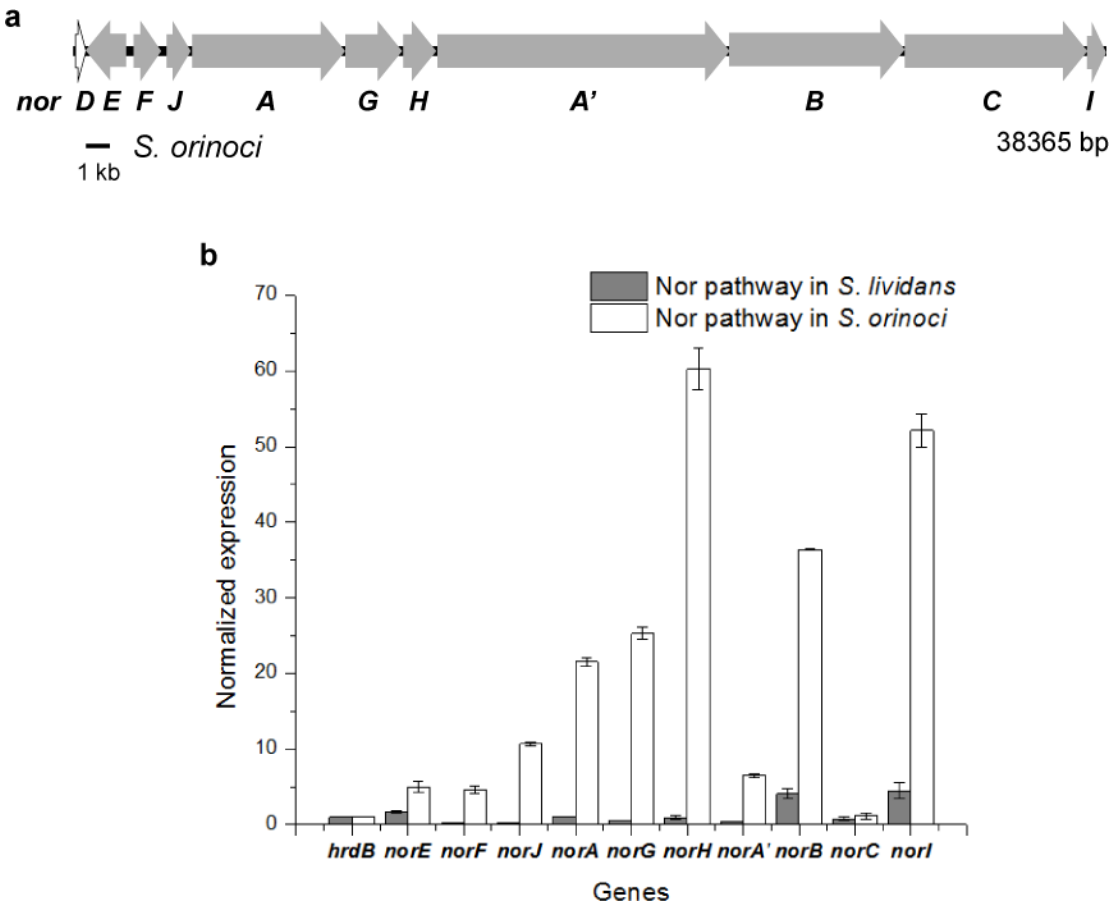


Figure 2.

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

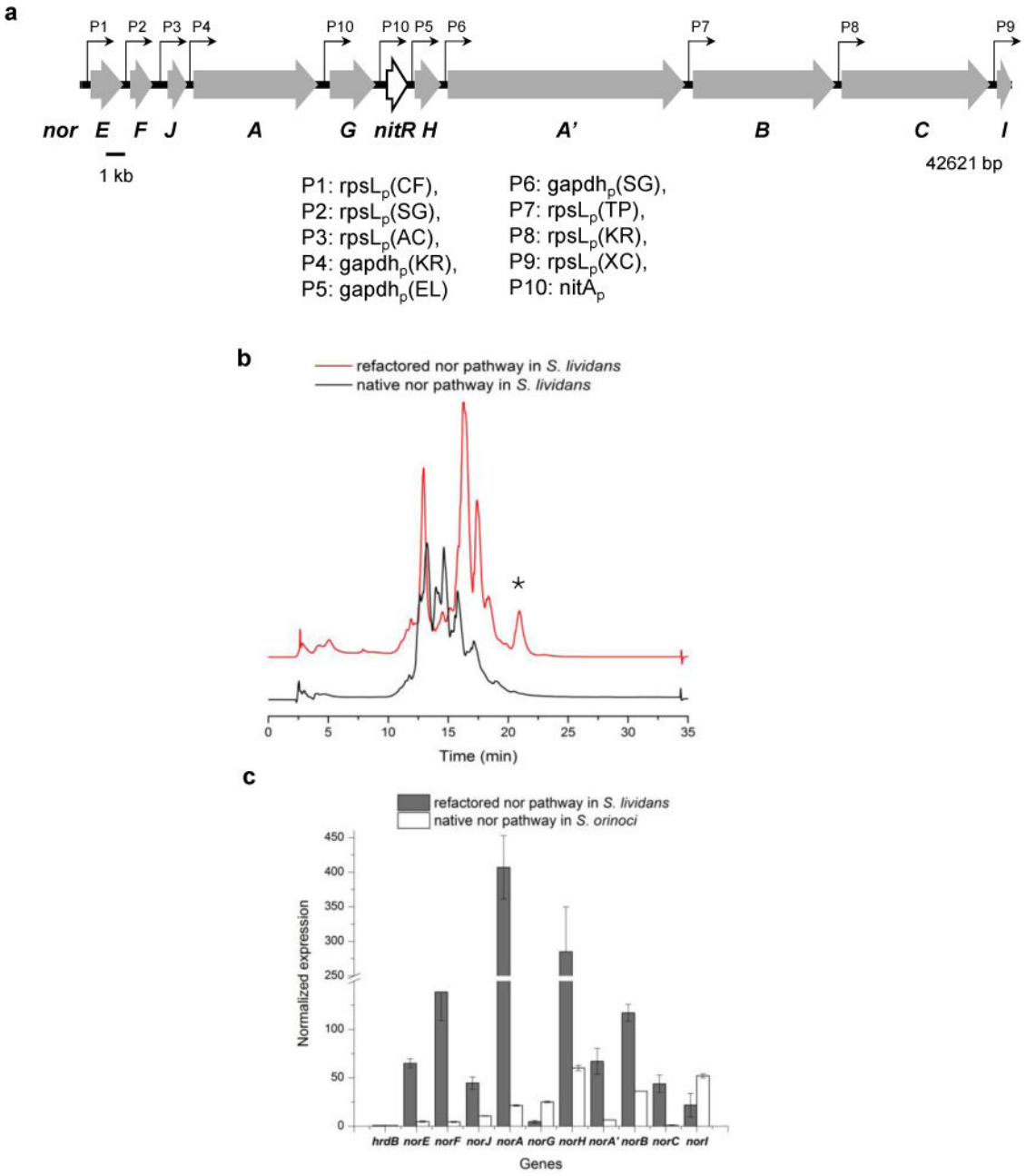
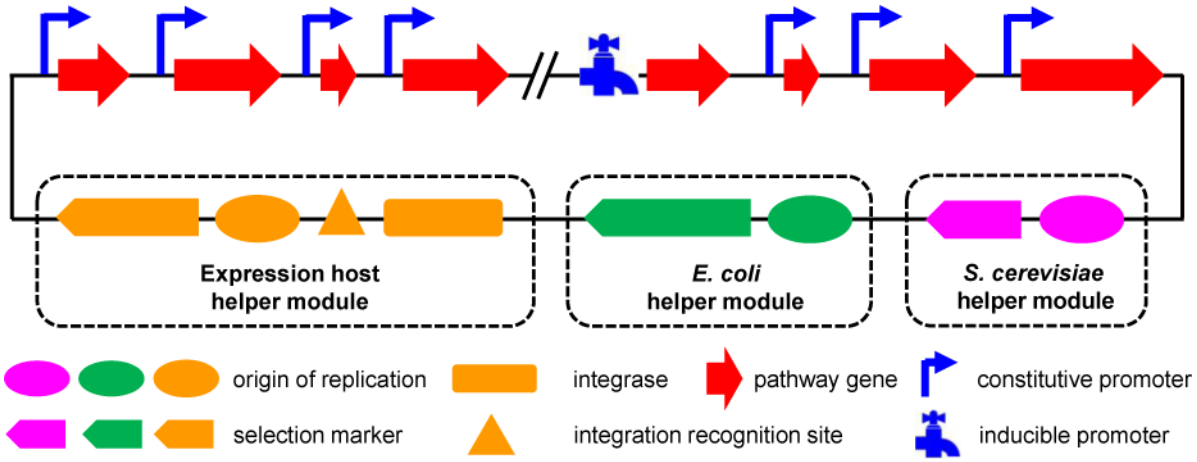


Figure 3.

a) The refactored spectinabilin pathway. b) LC-MS analysis of the extract from the *S. lividans* strain carrying the refactored *nor* pathway. The peak labeled by a star indicated the target product peak. c) Real-time PCR analysis of the *nor* gene transcription 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).



Scheme 1.

Design of a plug-and-play scaffold for refactoring cryptic natural product biosynthetic pathways. 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.