



HHS Public Access

Author manuscript

Nat Chem Biol. Author manuscript; available in PMC 2016 September 01.

Published in final edited form as:

Nat Chem Biol. 2015 September ; 11(9): 649–659. doi:10.1038/nchembio.1893.

Reinvigorating natural product combinatorial biosynthesis with synthetic biology

Eunji Kim¹, Bradley S. Moore^{2,*}, and Yeo Joon Yoon^{1,*}

¹Department of Chemistry and Nano Science, Ewha Womans University, Seoul, Republic of Korea

²Scripps Institution of Oceanography and Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, California, United States

Abstract

Natural products continue to play a pivotal role in drug discovery efforts and in understanding human health. The ability to extend nature's chemistry through combinatorial biosynthesis – altering functional groups, regiochemistry, and scaffold backbones through manipulation of biosynthetic enzymes – offers unique opportunities to create natural product analogues. Incorporating emerging synthetic biology techniques has the potential to further accelerate the refinement of combinatorial biosynthesis as a robust platform for the diversification of natural chemical drug leads. Two decades after the field originated, we discuss the current limitations, realities, and the state of the art of combinatorial biosynthesis, including the engineering of substrate specificity of biosynthetic enzymes and the development heterologous expression systems for biosynthetic pathways. We also propose a new perspective for the combinatorial biosynthesis of natural products that could reinvigorate drug discovery by using synthetic biology in combination with synthetic chemistry.

Chemists have long held a fascination with controlling and extending nature's biosynthetic dexterity. The emergence of combinatorial chemistry in the 1980s as a way to rapidly and efficiently produce massive chemical libraries that promised new drug leads inspired genetic engineers in the 1990s to follow suit. Their early efforts focused on reprogramming natural biosynthetic pathways by mixing and matching genes from known biosynthetic clusters to yield unnatural designer analogues of natural products differing by a single methyl group, the oxidation state of a C-O bond, or by more drastic changes to the scaffold itself. These studies, meant to fuse the capabilities of combinatorial chemistry with the genetic power and enzymatic prowess of biosynthesis, led to the concept of combinatorial biosynthesis¹.

Motivation for the field was high, as researchers anticipated numerous “unnatural” natural products with altered structures could be produced, illuminating structure-activity relationships key for drug development purposes and improving pharmaceutical properties of clinically relevant compounds. Even today, the importance of natural products as drug

*Correspondence should be addressed to B.S.M (bsmoore@ucsd.edu) or Y.J.Y. (joonyoon@ewha.ac.kr).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

leads cannot be overstated. They continue to account for the majority of antimicrobial and anticancer agents approved by the FDA² while similarly serving as indispensable molecular probes in illuminating fundamental cellular processes³. Nevertheless, the pharmaceutical industry has severely pulled back its support for natural products research in recent decades in favor of synthetic approaches in building chemical libraries that are more suitable for modern biological screening campaigns. Natural products, however, have maintained a several fold higher ‘hit rate’ than synthetic chemical libraries⁴ and continue to cover a much wider chemical space landscape⁵.

By exploiting the substrate promiscuity of biosynthetic machinery, researchers indeed were able to create new compounds with altered structures. For example, spinetoram is a spinosyn-based insecticide that was successfully developed from natural product derivatives through a combination of biological and chemical approaches. Marketed in 2007 with improved efficacy and an expanded spectrum⁶, it is a mixture of 3'-*O*-ethyl-5,6-dihydro-spinosyn J (**1**) and 3'-*O*-ethyl spinosyn L (**2**) (Fig. 1) synthesized by the chemical modification of spinosyns J and L produced by a rhamnose 3'-*O*-methyltransferase-inactivated *Saccharopolyspora spinosa*⁶. Although spinetoram was not developed solely by combinatorial biosynthesis, this successful case strongly suggests that more potent natural product analogues can be discovered by combinatorial biosynthesis. Yet limited knowledge about biosynthetic enzymes and pathways, such as the important roles played by the seemingly inert linker regions between enzymatic modules that were so often the site of genetic cutting and pasting, limited the scope of success, and even today we remain far from the envisioned future in which any desired small molecule can be programmed at will. We suggest this gap is due to still insufficient information about how individual biosynthetic enzymes function, how linked enzymatic modules function together, and limitations in tools to rapidly test relevant hypotheses. In this Perspective, we highlight both successes in the field as well as identify open questions that are hindering further research and point to growing synergies with synthetic biology that are galvanizing the field.

In the beginning

Early bioengineering success was achieved with microbial polyketides (PKs) and nonribosomal peptides (NRPs), compounds with important pharmaceutical activities such as antibacterial, anticancer and immunosuppressant^{7,8}. Their respective biosynthetic systems utilize simple malonate and amino acid building blocks to construct complex chemical structures in an assembly line fashion whereby intermediates are successively built while transiently tethered to carrier proteins. Classical combinatorial biosynthetic examples include the macrolide antibiotic erythromycin from a type I modular PK synthase (PKS)⁹, the aromatic PK actinorhodin from a type II PKS¹⁰, and the lipopeptide antibiotic daptomycin from a NRP synthetase (NRPS)^{11,12}. Hundreds of biosynthetic products (e.g. compounds **3** – **5**) have been engineered in these systems by manipulating substrate input and enzymatic assembly line context using a variety of strategies, including gene fusions, gene inactivations, gene replacements, enzyme domain substitutions, and module exchanges^{11–14} (Fig. 2). In the case of daptomycin, for instance, the combined use of these approaches have generated over 120 novel lipopeptides (e.g. **5**), some displaying improved therapeutic properties^{11,12} (Fig. 2c).

Expanding enzymatic diversity

Structural diversification by combinatorial biosynthesis was initially restricted by the substrate specificity of biosynthetic enzymes. However, in recent years, the biosynthetic community has excelled in characterizing new enzymes and altering the properties of known biosynthetic enzymes, both of which have important implications for combinatorial biosynthesis.

Newly characterized enzymes tend to fall into three camps – those responsible for constructing substrates, those for assembling substrates into the main structure, and those involved in product diversification. In the case of modular assembly line PKS and NRPS machines, progress has been made in all three areas. These include new substrate pathways to assemble and incorporate α -substituted malonyl thioesters in PKSs¹⁵ or nonproteinogenic amino acids in NRPSs¹⁶. Additional characterization of known types of tailoring enzymes that decorate these scaffolds, such as deoxyhexoses for modifying diverse biosynthetic aglycones¹⁷, as well as the discovery of new tailoring enzymes, such as those responsible for simple to exotic redox, halogenation, or alkylation reactions, offer new excitement for the bioengineering toolbox^{16,17}. Unfortunately, the details of tailoring enzymes are beyond the scope of this perspective article. Modifying known proteins through rational design, directed evolution or other engineering approaches, has also enabled the expansion or alteration of substrate specificity.

For example, the gatekeeper enzyme domain in modular PKSs is the acyltransferase (AT) domain that controls the selection and incorporation of monomeric building blocks or extender units (usually malonyl-, methylmalonyl-, or ethylmalonyl-CoAs) for the construction of PK products. The restricted versatility of the PK extender unit has limited the generation of novel PK structures with diverse side chains. However, over the past several years, the PKS substrate code has expanded significantly with the discovery of the crotonyl-CoA carboxylase/reductase (CCR) family of enzymes that catalyze the reductive carboxylation of α , β -unsaturated acyl-CoA precursors¹⁸. This expansion to rare acyl-CoA extender units such as haloethylmalonyl-CoA, allylmalonyl-CoA, isobutyrylmalonyl-CoA, and benzylmalonyl-CoA extends the structural possibilities of adding bulky or reactive side chains to the PK backbone¹⁵. Often AT domains associated with CCR-generated extender units display flexible extender unit-specificity, thereby providing opportunities to create diverse PKs with altered side chains. The AT domain of module 4 in the immunosuppressant FK506 PKS naturally accepts methylmalonyl-, ethylmalonyl-, propylmalonyl-, and allylmalonyl-CoA substrates as well as unnatural acyl-CoAs derived from supplemented carboxylic acids that generate macrolide derivatives with modified C21 side chains (compounds **6** – **8**). One of these novel compounds exhibited improved *in vitro* nerve regenerative activity relative to the parent FK506 (Fig. 1)¹⁹. Similarly the AT domain in AntD of the antimycin NRPS-PKS hybrid assembly line showed remarkable tolerance towards a variety of non-natural extender units, including the engineered and “clickable” 5-hexynylmalonyl-CoA²⁰ as well as linear, branched, cyclic and halogenated extender units²¹. This tolerance was exploited to produce diverse antimycin analogues (**9** – **12**) exhibiting enhanced *in vitro* cytotoxicity and antifungal activity (Fig. 1). Interestingly, the *de novo* synthesis of 5-hexynylmalonyl-CoA from three jamaicamide biosynthetic enzymes acting in

coordination with the promiscuous antimycin CCR has been employed to produce terminal alkyne-labeled natural products²⁰.

The incorporation of unnatural extender unit substrates however requires specific partner ATs to allow for their selective attachment to the modular assembly line. While some ATs exhibit broad substrate scopes, most are specific and narrow. A minimally invasive strategy to alter AT specificity while maintaining protein integrity involves amino acid substitutions. To illustrate, molecular docking simulation of the erythromycin PKS module 6 AT with methylmalonyl-CoA docked into the active site helped identify a residue that appeared to determine extender unit side chain size. Mutation of this residue (Val295Ala) was subsequently shown to incorporate the exogenously added non-natural extender unit 2-propargylmalonyl-SNAC (*N*-acetylcysteamine thioester) to produce 2-propargylerythromycin A (**13**)²² (Fig. 1). The expanded availability of new extender units derived from promiscuous CCRs¹⁵ or malonyl-CoA synthetase (MatB)²³ with naturally or engineered promiscuous AT domains provide great potential for the biosynthesis of novel PKs with specifically altered side chains.

In NRPS systems, the recognition of new nonproteinogenic amino acid substrates has ballooned to hundreds of diverse structures. The modification of the native adenylation (A) domains that control the entry of diverse amino acid substrates to the NRPS assembly line has proven most successful in engineering the selective attachment of unnatural substrates. For instance, the specificity of the A domain of module 10 within the calcium-dependent antibiotic (CDA) NRPS was modified by the introduction of a single mutation (Lys278Gln), changing its specificity from (2*S*,3*R*)-3-methyl Glu (mGlu)/Glu to (2*S*,3*R*)-3-methyl Gln (mGln)/Gln to produce the Gln/mGln-containing CDA analogues (**14** – **15**) (Fig. 1)²⁴. A similar point mutation in the Phe-specific A domain of the gramicidin synthetase GrsA (Trp239Ser) allowed for the introduction of an *O*-propargyl-Tyr residue²⁵.

In addition to this rational design approach in NRPS engineering, directed evolution has been successfully applied to the reprogramming of A domain specificity. For instance, the Val-activating A domain of AdmK within the hybrid andrimid NRPS-PKS pathway was engineered. Three of the most highly variant residues were selected among the specificity-conferring sites for saturation mutagenesis, and the resulting library created in the native producer, *Pantoea agglomerans*, was screened for the production of analogues by LC-MS analysis of the crude cell extract. As a result, four clones were isolated to produce andrimid derivatives, three of which (**16** – **18**) (Fig. 1) showed increased bioactivity against *Staphylococcus aureus*²⁶. A key aspect of this study is the structure-based assay in the native producing host that permits a direct read-out of not only the engineered AdmK A domain's activity but also the combined output of downstream biosynthetic enzymes. However, andrimid system is a relatively small NRPS and the expandability of this structure-based screening method to multimodular NRPSs or PKSs should be tested further. In addition, this process requires a great deal of screening effort using LC-MS, and over ~ 14,000 clones of the library were screened to isolate the four clones producing andrimid analogues.

Building up new machines

These examples illustrate a few case studies in which unnatural building blocks can be selectively incorporated into PKs or NRPs using non-invasively engineered AT or A domains. An alternate approach is to replace native domains with domains from other pathways that already contain the desired specificity. For example, orthogonal *trans*-AT PKSs, in which the AT domain is rather located on a freestanding enzyme instead of on the multimodular synthase²⁷, provide one successful case study of domain swapping: these domains are typically dedicated to malonyl-CoA. However, the *trans*-AT protein KirCII from the kirromycin PKS is naturally specific for ethylmalonyl-CoA and possess a relaxed specificity towards non-natural extender units *in vitro*²³. Notably, the *trans*-acting AT from the disorazole PKS allowed for the efficient incorporation of fluoromalonyl-CoA in a *cis*-AT-inactivated unimodular or bimodular erythromycin PKS model system to produce novel fluorinated PKs such as 2-methyl-4-fluoro-tetraketide lactone (**19**) and 2-fluoro-4-methyl-tetraketide lactone (**20**) (Fig. 1)²⁸.

While this example provides one success story of AT domain replacements, this strategy typically results in impaired protein folding or interactions that often significantly reduce biosynthetic productivity²⁹. Efforts to create new NRP analogues by swapping A domains similarly result in reduced product yields when compared to native systems^{11,12,16}.

Although it has been demonstrated that directed evolution can improve the productivity of an engineered NRPS replaced with a heterologous A domain³⁰, maintaining the functional integrity of these megasynthases during domain replacements or insertions remains a substantial challenge given our limited understanding of the structural and temporal restrictions of this complex machinery; i.e., the mode, timing, and consequences of the inter- and intramolecular interactions of megasynthases. Recent structural studies of NRPS³¹ and PKS assembly lines^{32,33} have provided insight into some of the intramolecular contacts and motions of these proteins (see page xxx). A recently developed thiocyanate vibrational spectroscopic probe installed on the terminal thiol of the ACP's 4'-phosphopantetheine arm³⁴, which enables the observation of ACP conformational changes during biosynthesis, will also be a useful tool to comprehend PKS operating principles and to design productive hybrid PKSs. Although developing universal rules to bring together disparate enzymatic parts and design new products is still a major challenge, this provides the opportunity to utilize new basic knowledge to develop engineering strategies.

Another area of assembly line construction that is poorly understood involves iterative versus single-use modular biochemistry. Most modular PKS and NRPS proteins construct their products via a linear assembly line process where each and every enzymatic domain is utilized once per assembled product^{7,8}. In some cases, however, enzymatic domains or modules are entirely skipped or used more than once resulting in diversity elaboration. One of the more remarkable multitasking synthetases is the hybrid PKS-NRPS associated with the thalassospiramide family of calpain inhibitors from *Thalassospira* marine α -proteobacteria, where 16 lipocyclic peptides result from an assortment of amino acid substrates being channeled through enzymatic multimodule skipping and iteration reactions³⁵. This extreme natural example reveals that assembly line systems are amenable for library construction, yet the programming rules are not understood.

Recently, a ratchet mechanism was proposed that ensures the unidirectional transport of a growing PK chain along the PKS³⁶. Based on this model, module 3 of the erythromycin PKS was reprogrammed to catalyze two rounds of chain elongation in succession by employing an engineered ACP domain, in which the N terminus region that principally interacts upstream from the ketosynthase (KS)-AT didomain was replaced with the corresponding sequence from ACP domain of module 2 to permit back-transfer of the chain elongation product that initially forms to the KS domain of module 3³⁶. Additionally, the corresponding PKS and NRPS subunits must interact with their appropriate partners, and these interactions are mediated in part by docking domains^{37,38} and communication-mediating domains³⁹, respectively, at the termini of the subunits. Using the recently characterized docking domains from the cyanobacterial curacin A PKS, it was demonstrated that the selectivity of docking interactions could be altered by site-directed mutagenesis⁴⁰. These examples provide compelling case studies demonstrating that the combination of basic and applied science can enable future efforts in the fine reprogramming of hybrid megasynthase.

Synthetic biology to the rescue

Until recently, one of the major impediments to combinatorial biosynthesis was the lack of available DNA sequences for natural product biosynthetic genes. Now that advanced sequencing technologies have created an enormous pool of natural product biosynthetic genes, new methods for the rapid assembly of large gene clusters are needed^{41–43}. Multidisciplinary synthetic biology approaches are poised to enable combinatorial biosynthesis studies to be faster and more accurate^{16,44} by providing genetic tools to enable the rapid assembly and refactoring of large biosynthetic genes and by providing a series of optimized hosts or “chassis” for the heterologous expression of diverse biosynthetic pathways. Specifically, a variety of newly available ligation-based and homology-based synthetic biology techniques have now been developed for DNA assembly^{45,46}. Homology-based methods have been extensively utilized for the heterologous expression of natural product gene clusters, and studies where these techniques have been successfully used to produce natural products are highlighted below.

Conventional Red/ET recombineering is an *in vivo* phage-mediated homologous recombination based method used in *Escherichia coli* between linear and replicating circular DNA (linear plus circular homologous recombination, LCHR) catalyzed by the Red $\alpha\beta\gamma$ from the coliphage λ Red system or a truncated version of RecE along with RecT from the Rac prophage (Fig. 3)⁴⁷. LCHR was successfully used to exchange single or multidomains for the combinatorial biosynthesis of the daptomycin derivatives described above^{11,12}. Over the past decade, a range of natural product gene clusters has been heterologously expressed by this method⁴¹. In spite of this success, Red/ET recombineering becomes difficult when assembling large clusters cloned on multiple overlapping clones since step-wise recombination with a unique selectable marker is required. More recently, it was discovered that full-length RecE with RecT efficiently facilitates linear plus linear homologous recombination (LLHR) between two linear DNA molecules with a more than 20-fold greater efficiency when compared with LCHR (Fig. 3)⁴⁸. LLHR was successfully utilized for the direct cloning of ten unknown PKS/NRPS gene clusters from *Photorehabdus luminescens*

into *E. coli* circumventing library construction, step-wise stitching, and screening. Two of these gene clusters were successfully expressed in *E. coli* and produced the compounds designated as luminmide A (**21**) and luminmycin A (**22**) (Fig. 4)⁴⁸. Although LLHR allows for the quick and efficient direct cloning of a target DNA sequence from a genomic DNA template, the ultimate efficiency of this approach is limited by the size of the target cluster, and two steps of LLHR would be required for the direct cloning of gene clusters larger than 60 kb.

In addition to these phage-mediated homologous recombination methods, transformation-associated recombination (TAR)-based techniques that take advantage of the native *in vivo* homologous recombination of *Saccharomyces cerevisiae*⁴⁹ have been developed to directly capture gene clusters from environmental DNA samples or large genomic loci (Fig. 3). By utilizing a combination of environmental DNA library screening and TAR to assemble the biosynthetic machinery on the bacterial artificial chromosome (BAC), the novel natural products fluostatin F, G, and H (**23** – **25**)⁵⁰ and pentacyclic ring (**26**)⁵¹ were successfully produced in *Streptomyces albus* (Fig. 4). More recently, a cosmid-based TAR system has been developed. This *S. cerevisiae*/*E. coli* shuttle-actinobacterial chromosome integrative capture vector (named pCAP01) can be kept at multiple copies in *E. coli* due to the pUC ori derived from cosmid. This supplies an ample amount of plasmid DNA without induction, in contrast to BAC-based TAR systems. With this method, a silent 67-kb NRPS gene cluster from the marine actinomycete *Saccharomonospora* sp. CNQ-490 was captured, activated and expressed in *S. coelicolor* yielding a dihalogenated derivative of daptomycin, taromycin A (**27**)⁵² (Fig. 4). This direct cloning approach, which allows for a single genomic capture and expression vector, is compatible with next generation technology that does not require large insert clonal libraries. These TAR-based methods are compatible with Red/ET recombination-mediated PCR targeting⁵³ for gene replacement to support subsequent engineering of the assembled gene cluster.

Another TAR-based method, “DNA assembler”, enables convenient gene deletions and replacements. With DNA assembler, genetic fragments encoding the biosynthetic pathway of interest and required for DNA replication in *S. cerevisiae*, *E. coli*, and other heterologous hosts are designed to overlap with two adjacent fragments amplified from the native producer genome or the corresponding vectors; these are subsequently co-transformed into *S. cerevisiae* for homologous recombination (Fig. 3). Recently, the ~29 kb aureothin pathway was assembled by seven 4 – 5 kb fragments using the DNA assembler method and was expressed in *Streptomyces lividans* producing aureothin⁵⁴. In addition, since this method only requires adding site-specific mutations into the PCR primers for site-directed mutagenesis, the DH domain of the AurB PKS was readily inactivated to generate aureothin analogue (**28**)⁵⁴ (Fig. 4). However, with the DNA assembler method, random mutations are likely to occur due to the multiple PCR amplification steps. Moreover, mutations, deletions or insertions can be introduced if fragments pair incorrectly during the assembly process, particularly when PKS or NRPS domains show high sequence identity.

One of the best known *in vitro* homology-based methods is the one-pot isothermal assembly known as “Gibson assembly”⁵⁵, in which DNA fragments with overlapping sequences are joined via three enzymatic reactions. First, T5 exonuclease degrades the 5' ends of each

fragment to expose their complementary single-stranded 3' ends, which then anneal to each other in the designed order with single-stranded gaps, that are filled by Phusion polymerase and then the nicks are sealed by Taq ligase to produce the final product. Sequence and ligation independent cloning (SLIC)⁵⁶ is one variation of this *in vitro* homology-based method, where T4 DNA polymerase is utilized for both 3' single-strand degradation and partial gap-filling (Fig. 3). SLIC has been applied to the combinatorial biosynthesis of premonensin derivatives⁵⁷. Mutated fragments of β -carbon processing domains of type I monensin PKS were rapidly cloned into the temperature-sensitive shuttle vector using SLIC and were introduced through in-place homologous recombination into the chromosomal monensin PKS genes of *Streptomyces cinnamonensis* A495 leading to a library of 22 premonensin redox derivatives with a few (**29** – **31**) (Fig. 4) showing increased antibacterial activity against *Pseudomonas aeruginosa*⁵⁷. Although the construct size that can be assembled *in vitro* seem to be smaller than *in vivo*, a distinct advantage of *in vitro* DNA assembly is that it can be accomplished in hours, while *in vivo* methods can take days.

Toward optimal chassis for heterologous biosynthesis

Sequences from diverse bacteria, such as myxobacteria and cyanobacteria⁴¹, fungi⁴¹, plants⁵⁸ and even environmental DNA⁴² are now increasingly available in biosynthetic engineering experiments. With the development of diverse and facile DNA assembly techniques, robust chassis strains are also needed for the successful heterologous production of new natural products using this remarkably increased and publically available (meta)genomic sequence information and further combinatorial biosynthesis.

Currently, the choice of a suitable chassis is often determined by the source of the target pathway and metabolite type. Although there is no universal chassis available, the most widely used microbial chassis are *Streptomyces* species and *E. coli*. *Streptomyces* species have several advantages. First, as this bacterial genus supports most biosynthetic pathways, secondary metabolic biosynthetic precursors are readily available from primary metabolism. Second, as *Streptomyces* are physiologically different from *E. coli* (e.g. high GC content and Gram(+)), their proteins, promoters, activators and regulatory processes are often not compatible with those in *E. coli*. And third, *Streptomyces* support distinct post-translational modification enzymes (e.g. phosphopantetheinyl transferase (PPTase) responsible for adding a 4'-phosphopantetheine group to PKS and NRPS carrier proteins) required for functional biosynthetic enzymes. While many different *Streptomyces* hosts have been used for the expression of heterologous genes, *S. coelicolor*, *S. lividans*, *S. albus*, and the avermectin producer *Streptomyces avermitilis* are most commonly employed. Removal of endogenous secondary metabolite genes from the chassis strain is advantageous because precursors and energy are redirected for the production of the heterologous compound, and a cleaner metabolite background simplifies the detection and purification of the heterologous compound. To this end, a derivative of *S. coelicolor*⁵⁹ and *S. avermitilis*⁶⁰ were constructed, in which active biosynthetic gene clusters were eliminated. The feasibility of these mutants as efficient chassis was shown by the successful production of a variety of natural products derived from other actinomycetes and plants^{59–62}.

Another efficient strategy is to use industrial overproduction strains that have already been optimized by conventional mutation. The overproduction properties of the ‘clean-host’ derivative of *Saccharopolyspora erythraea* industrial strain were exploited to enhance titers of heterologous PKs. An approximately 10-fold improved production of erythromycin analogues (**32** and **33**) (Fig. 4) were observed in the engineered *S. erythraea* overproducer by chromosomal integration of engineered PKs compared to *S. lividans*⁶³. This suggested that the characteristics of the overproducing strain are primarily due to mutations in non-PKS genes and thus, operate equally well on other PKs. Multi-gram quantities of tetracenomycins were produced by the heterologous expression of type II tetracenomycin PKS genes from *Streptomyces glaucescens* in a monensin A-high producing *S. cinnamonensis*⁶⁴. Interestingly, the deletion of the monensin biosynthetic genes did not affect the production of tetracenomycin.

Despite these successful examples of actinomycete chassis, *E. coli* still offers certain advantages in terms of a much faster growth rate and a variety of facile genetic tools. One of the disadvantages of *E. coli* as a chassis is that not all of the biosynthetic machinery, such as PPTase, or substrates, such as methylmalonyl-CoA, are available. The first successful production of a type I PK product, 6-deoxyerythronolide B (the PK aglycone core of erythromycin A), was achieved using an *E. coli* host (termed BAP1) engineered to be capable of posttranslationally modifying PKS ACP domains and to supply methylmalonyl-CoA in sufficient amounts⁶⁵. More recently, reconstitution of the entire erythromycin pathway in *E. coli* was attempted; erythromycin A was successfully heterologously produced in *E. coli* using six plasmids harboring three large PKS genes and 17 additional genes responsible for deoxysugar biosynthesis, macrolide tailoring, and resistance⁶⁶. This *E. coli* platform was further applied to the production of erythromycin analogues (**34** and **35**) (Fig. 4) by replacing the loading module and altering the extender unit specificity of the second module, respectively⁶⁶. NRPs such as echinomycin from *Streptomyces lasaliensis*⁶⁷, syringolin derivatives derived from *Pseudomonas syringae*⁶⁸, and alterochromide lipopeptides from *Pseudoalteromonas piscicida* JCM 20779⁶⁹, as well as the NRP-PK hybrid antimycin from *S. ablus* or *Streptomyces ambofaciens*⁷⁰ were also produced in *E. coli*. These results demonstrate *E. coli* is undoubtedly an interesting alternative chassis for heterologous expression and further combinatorial biosynthesis.

In addition to *Streptomyces* and *E. coli*, other hosts such as *Pseudomonas putida*, *Myxococcus xanthus*, and *S. cerevisiae* have also been widely used for the heterologous production of natural products⁴¹. It is worth noting that *S. cerevisiae* is a viable chassis for the heterologous production of plant and fungal metabolites. Reconstitution of plant pathways in yeast has been proved to be a promising approach to access plant metabolites such as the terpenoid antimalarial drug precursor artemisinic acid⁷¹ and strictosidine, the last common biosynthetic intermediate for all monoterpene indole alkaloids⁷². The iterative fungal PKS LovB from the lovastatin pathway of *Aspergillus terreus* was also successfully reconstituted in *S. cerevisiae*⁷³. Recently, another efficient system for the heterologous expression of fungal gene clusters in *Aspergillus nidulans* was developed. The asperfuranone PKS gene cluster from *A. terreus* was assembled using fusion PCR and *in vivo* homologous recombination during transformation under the controllable *alcA* promoter

in *A. nidulans* where its own cryptic asperfuranone cluster was deleted. The resulting *A. nidulans* mutant produced ~ 7 mg/L of asperfuranone⁷⁴. It is possible that this system may work for other clusters of different metabolite types or derived from fungi other than *Aspergillus*, thus providing a general system for fungal metabolite production.

Despite these advances, a single all-purpose chassis capable of producing large quantities of diverse natural products has not yet been developed. Perhaps that is an unrealistic expectation. Rather, it is more realistic to envision the development of a set of genetically tractable, genome-minimized hosts optimized for each metabolite type capable of supplying the precursors for the targeted secondary metabolites and efficiently expressing the target biosynthetic pathways in a controllable manner (e.g. *Streptomyces* species for large microbial PKSs or NRPSs; *E. coli* for small PKSs, ribosomally synthesized and post-translationally modified peptides (RIPPs), or other small enzymes; and yeast for plant and fungal enzymes).

Potential synthetic biology tools for biosynthesis

Although most of the currently developed synthetic biology tools are designed for *E. coli* and may not be directly applicable for combinatorial biosynthesis in actinomycetes, advancements in synthetic biology methodology offer new approaches to address the problems associated with combinatorial biosynthesis (Fig. 5 – 6). For example, it is now possible to employ synthetic DNA to functionally reconstitute large PKS-NRPS pathways, thereby enabling the use of optimized codons in desirable hosts (Fig. 5). The promising family of anticancer drugs, the epothilones, is biosynthesized by a hybrid NRPS-PKS in the myxobacterium *Sorangium cellulosum*, which is notorious for its difficult genetic manipulation and very slow growth rate. Therefore, attempts have been made to express the codon-optimized synthetic epothilone gene cluster into faster growing and more genetically tractable chassis. To this end, the entire ~55 kb GC-rich epothilone gene cluster was codon-optimized to ensure efficient translation, redesigned to place a strategic restriction site, and synthesized to express these large proteins in *E. coli*. The expression of the largest tetramodular EpoD (765 kDa) containing PKS modules 3 – 6 required splitting this protein into two bimodular polypeptides with compatible docking domains to maintain the interaction between EpoD module 4 and module 5. The combined use of an alternative promoter, lowered temperature, and coexpression of chaperones led to the production of soluble proteins from all epothilone biosynthetic genes and the subsequent production of epothilones C and D at <1 µg/L in a strain of *E. coli* BAP1⁷⁵. Similarly, the functionality of the codon-optimized artificial pathway was demonstrated by the heterologous production of epothilones A and B (~ 100 µg/L) in *M. xanthus*⁷⁶. Although the reconstitution of large PKS/NRPS biosynthetic pathways using synthetic genes has so far only been demonstrated for the epothilones, the achieved low production yields suggest that unknown fundamental constraints might limit the utilization of artificial DNA for the production of certain classes of PK or NRP natural products.

Additionally, there are opportunities to engineer the spatial organization of biosynthetic enzymes. Synthetic scaffolds based on protein, RNA, and DNA have been successfully engineered to co-localize proteins and to increase the yield of the biosynthetic pathway by

Author Manuscript

Author Manuscript

Author Manuscript

increasing the spatial proximity of biosynthetic enzymes, optimizing their local stoichiometry, and minimizing the loss of intermediates while avoiding accumulation of toxic intermediates (Fig. 5). In particular, a protein scaffold built using the protein-protein interaction domains of signaling proteins to recruit metabolic enzymes fused with their specific peptide ligands was used to co-localize and control the stoichiometry of three mevalonate biosynthesis enzymes. Significant (77-fold) increases in mevalonate production were observed by introducing several of these scaffolds in *E. coli*⁷⁷. Alternatively, RNA aptamer-based scaffolds, in which RNA aptamers incorporated into the scaffolds recruited the appropriately tagged enzymes to precise positions on the scaffold, have been reported to control the spatial organization of two enzymes involved in hydrogen production, yielding a 48-fold increase in hydrogen production over a system without the scaffold⁷⁸. More recently, DNA scaffolds built by creating chimeras between biosynthetic enzymes and zinc finger domains that bind DNA sequences in a controlled order were used to moderate titer enhancement (up to 5-fold) of resveratrol, 1,2-propanediol, and mevalonate in *E. coli*⁷⁹. Despite the potential disadvantages of these synthetic scaffold systems, such as the instability of the scaffolds, these examples suggest the potential use of artificial scaffold systems to efficiently co-localize and control the stoichiometry of individual enzymes in non-PKS or NRPS pathways.

Moreover, promising fine-tunable libraries of promoters⁸⁰ and ribosome binding sites (RBSs)⁸¹ can control the relative amounts of individual enzymes at both the transcriptional and translational levels in *E. coli*. Recently, a synthetic promoter library for actinomycetes based on the randomization of the –10 and –35 consensus sequence of the widely used constitutive promoter ermEp 1 from *S. erythraea* was also constructed displaying 2–319% of ermEp 1 activity. The strongest synthetic promoter was used to express a type III PKS RppA from *S. erythraea* and an approximately 3-fold increase in flaviolin production was observed compared with the ermEp 1* promoter (a strong variant of ermEp 1) in *S. lividans* (Fig. 5)⁸².

High-throughput genome editing tools, such as the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) protein (CRISPR/Cas) system⁸³, and genomic evolution technology, such as the multiplex automated genome engineering (MAGE)⁸⁴, are now available for the efficient engineering of microbial hosts for heterologous expression and combinatorial biosynthesis via chromosomal deletion, the addition or alteration of the specific sequences or the fine-tuning of expression level of multiple targeted genes (Fig. 6). In the type II CRISPR/Cas system, which provides bacterial immunity against foreign DNA via RNA-guided DNA cleavage, short foreign ‘spacer’ DNAs are integrated within the CRISPR genomic loci and transcribed into short CRISPR RNAs (crRNA) which anneal to trans-activating crRNAs (tracrRNAs) to direct sequence-specific cleavage of foreign DNA by Cas9 proteins. Recent evidence suggests that the Cas9/crRNA/tracrRNA complex can target any protospacer DNA sequence when the requisite NGG protospacer-adjacent motif (PAM) is present at the 3' end. The spacer sequences matching target genomic loci can be directly programmed into a heterologously expressed CRISPR array, and the crRNA and tracrRNA can be fused to create a chimeric, single-guide RNA (sgRNA) for further simplification. The CRISPR/Cas system can be reconstituted in a

range of hosts including human cells⁸⁵, *S. cerevisiae*⁸⁶, *E. coli*⁸⁷, and in different *Streptomyces* species⁸⁸, demonstrating the potential for multiplex genome editing (Fig. 6). However, further studies are needed to assess its utility, including the extent to which off-target effects occur. Alternatively, MAGE is an efficient method for generating targeted diversity on a genome scale to achieve the coordinated expression of individual genes for maximum efficiency of the host system. Single-stranded oligonucleotides can be used to target multiple short DNA for directed evolution through mismatches, insertion, and deletion in the *E. coli* genome. The oligonucleotides can be sequentially transformed into the same *E. coli* population allowing the targeted modification to accumulate at different chromosome sites. The efficiency of MAGE was demonstrated by isolating an *E. coli* variant with a more than 5-fold increase in lycopene production within 3 days (Fig. 6)⁸⁴. However, because this technology requires an efficient transformation and single-stranded oligonucleotide-mediated allelic replacement, it may not be amenable for wider use in other organisms and has not been adapted for use in actinomycetes with a high GC content.

Additionally, antisense RNA (asRNA)-mediated gene silencing is also an effective strategy to engineer the host without genome editing or gene knockouts that can cause host instability or may result in decreased growth rates. This approach has been used in *E. coli* and also in other bacteria including *S. coelicolor* where antisense RNA silencers successfully inhibited actinorhodin production (Fig. 6)⁸⁹. In addition, this approach has recently been used with small regulatory RNAs (sRNAs), short noncoding RNAs in prokaryotes that control target gene expression at the post-transcriptional level, thus controlling gene expression in *E. coli* (Fig. 6). Engineered synthetic sRNAs consist of a target-binding sequence complementary to the target mRNA and a scaffold for recruiting the Hfq protein, which helps the hybridization of sRNA to the target mRNA and the subsequent degradation. Using this strategy to combinatorially knock down candidate genes in different strains, an *E. coli* strain was isolated that produces 2 g/L of tyrosine⁹⁰. This method is useful for large-scale target identification and for modulating chromosomal gene expression without modifying these genes, but unfortunately, the applicability to other hosts has not yet been shown.

When developed for use in a wider range of chassis including the most prolific natural product-producing actinomycetes, these synthetic biology tools listed above will facilitate essential combinatorial biosynthesis studies aimed at broadening the scope of biosynthetic building blocks/enzymes and developing efficient and robust heterologous expression systems. For instance, the spatial organization and stoichiometry of the enzymes expressed from the codon-optimized synthetic genes can be engineered using synthetic scaffolds in desirable hosts in which their precursor supply, regulatory networks, and other production properties are optimized by CRISPR/Cas, MAGE, or sRNA methods.

Perspective

To be useful for drug discovery, large chemical libraries need to be generated for screening, but in most applications, the combinatorial biosynthesis of natural products is not truly “combinatorial”. Rather, this process creates small focused libraries of natural products that can still reduce costs and time associated with drug discovery due to the significantly higher

hit rate for natural products and their derivatives⁴. Combining biosynthesis and synthetic organic chemistry allows combinatorial biosynthesis to be “more combinatorial” and provide greater promise in delivering chemical libraries for lead optimization and structure-activity relation studies. Recently, the multiplex combinatorial biosynthesis of a total 380 antimycin derivatives (e.g. **10** – **12**), of which 356 were new, was realized by combining multiple mutasyntheses at different biosynthetic stages with the chemistry needed to provide unnatural precursors and to modify the alkyne-bearing variants via Click chemistry²¹. In another example, the chlorine installed on pacidamycin by the expression of a heterologous halogenase was used as a selective handle for further synthetic diversification to generate four new analogues (**36** – **39**) (Fig. 4) using mild cross-coupling conditions in crude aqueous extracts of the culture broth⁹¹. Although conventional chemical methods for structural diversification will remain an indispensable technology, combinatorial biosynthesis reinvigorated by synthetic chemistry has now become an efficient alternative approach. Scientists in both fields should co-operate more closely to design useful analogues as well as to generate compounds that are difficult to be made solely by chemistry.

Although early success stories focused on PKs and NRPs primarily from actinomycete bacteria, combinatorial biosynthesis has since been applied to many classes of natural products, including RIPPs⁹², saccharides⁹³, terpenoids⁹⁴ and alkaloids⁹⁵ from assorted organisms. Because the non-modular biosynthetic machineries for these natural products are more amenable for combinatorial assembly than PKS and NRPS systems, they have huge potential to generate combinatorial biosynthetic libraries. For example, RIPP biosynthetic genes from cyanobacteria including the patellamide⁹⁶ and cyanobactin pathways⁹⁷ were functionally expressed in *E. coli*. In the latter case, rules for the sequence selectivity of a RIPP pathway were determined suggesting that RIPP pathways are a promising source for combinatorial engineering.

Once we know what to make, the challenge remains how to make it at scale. Despite many successful examples of combinatorial biosynthesis, the production of many natural products generated by combinatorial biosynthesis suffers from low productivity, ultimately delaying analysis in the lab and clinical testing or commercialization of compounds with improved activity or pharmacokinetics. We posit that improved yields will be realized through the integration of emerging systems and synthetic biology techniques combined with traditional metabolic engineering to optimize chassis strains and expression systems, as exemplified by the industry-scale production of the antimalarial drug precursor artemisinic acid in engineered yeast⁷¹.

With the increasing scientific understanding and technical capabilities gained over the last two decades, we argue that the basic foundation is laid for the “total biosynthesis” of designed natural products or small focused libraries solely through combinatorial biosynthesis. Further investigations of biosynthetic systems will guide our strategies, synthetic biology will enable the construction of naturally unavailable but efficient biological machinery, and synthetic chemistry will allow us to install naturally unavailable chemical functionalities to truly diversify nature’s biosynthetic chemistry. We envision that in the not-too-distant future, the structures of target molecules derived either from unisolable or uncultivable natural sources or by rational design will be able to guide the retro-synthetic

design of artificial biosynthetic genes using computational tools⁹⁸ (see page xxx) and synthesized in parts. The resulting synthetic DNA can then be assembled into functional biosynthetic pathways and productively expressed in an optimized chassis for the *de novo* generation of these designed products. The feasibility of this envisioned future is exemplified in the construction of the biosynthetic pathway of the antiviral nucleoside analogue 2',3'-dideoxyinosine (didanosine) in a retrograde fashion⁹⁹. As the production of the final product is the only selection criterion in this 'bioretrosynthesis' approach, assay design and the screening of each biosynthetic step can be reduced. After directed evolution of the final and penultimate enzymes, purine nucleoside phosphorylase and 1,5-phosphopentomutase, respectively, the antepenultimate enzyme ribokinase was engineered unexpectedly to bypass the 1,5-phosphopentomutase step due to the increased direct phosphorylation activity of the anomeric hydroxyl group of the starting substrate dideoxyribose. This shortened pathway showed a 50-fold increase in didanosine production *in vitro*⁹⁹.

The field of combinatorial biosynthesis is now ready to tackle the current challenges in natural product drug development.

Acknowledgments

We thank Dr Kris Rathwell for critically reading this manuscript. Research in Y.J.Y.'s laboratory has been supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (MISP) (2013R1A2A1A01014230), the Intelligent Synthetic Biology Center of the Global Frontier Project funded by MISP (20110031961), High Value-added Food Technology Development Program, Ministry of Agriculture, Food and Rural Affairs, Republic of Korea, and the National Research Council of Science and Technology through the Degree & Research Center program (DRC-14-3-KBSI). Combinatorial biosynthetic work in B.S.M.'s laboratory is supported by National Institutes of Health grants R01-CA127622 and R01-GM085770.

References

1. Hopwood DA, et al. Production of 'hybrid' antibiotics by genetic engineering. *Nature*. 1985; 314:642–644. [PubMed: 3857464]
2. Newman DJ, Cragg GM. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod*. 2012; 75:311–335. [PubMed: 22316239]
3. Robles O, Romo D. Chemo- and site-selective derivatizations of natural products enabling biological studies. *Nat Prod Rep*. 2014; 31:318–334. [PubMed: 24468713]
4. Li JW, Vederas JC. Drug discovery and natural products: end of an era or an endless frontier? *Science*. 2009; 325:161–165. [PubMed: 19589993]
5. Ertl P, Schuffenhauer A. Cheminformatics analysis of natural products: lessons from nature inspiring the design of new drugs. *Prog Drug Res*. 2008; 66:219–235.
6. Huang KX, Xia L, Zhang Y, Ding X, Zahn JA. Recent advances in the biochemistry of spinosyns. *Appl Microbiol Biotechnol*. 2009; 82:13–23. [PubMed: 19082588]
7. Hertweck C. The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl*. 2009; 48:4688–4716. [PubMed: 19514004]
8. Schwarzer D, Finking R, Marahiel MA. Nonribosomal peptides: from genes to products. *Nat Prod Rep*. 2003; 20:275–287. [PubMed: 12828367]
9. McDaniel R, et al. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc Natl Acad Sci USA*. 1999; 96:1846–1851. [PubMed: 10051557]
10. McDaniel R, Ebert-Khosla S, Hopwood DA, Khosla C. Engineered biosynthesis of novel polyketides. *Science*. 1993; 262:1546–1550. [PubMed: 8248802]

11. Nguyen KT, et al. Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc Natl Acad Sci USA*. 2006; 103:17462–17467. [PubMed: 17090667]
12. Baltz RH. Combinatorial biosynthesis of cyclic lipopeptide antibiotics: a model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways. *ACS Synth Biol*. 2014; 3:748–758. [PubMed: 23654258]
13. Sugimoto Y, Ding L, Ishida K, Hertweck C. Rational design of modular polyketide synthases: morphing the aureothin pathway into a luteoreticulins assembly line. *Angew Chem Int Ed Engl*. 2014; 53:1560–1564. [PubMed: 24402879]
14. Wang P, Kim W, Pickens LB, Gao X, Tang Y. Heterologous expression and manipulation of three tetracycline biosynthetic pathways. *Angew Chem Int Ed Engl*. 2012; 51:11136–11140. [PubMed: 23024027]
15. Wilson MC, Moore BS. Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Nat Prod Rep*. 2012; 29:72–86. [PubMed: 22124767]
16. Giessen TW, Marahiel MA. Ribosome-independent biosynthesis of biologically active peptides: Application of synthetic biology to generate structural diversity. *FEBS Lett*. 2012; 586:2065–2075. [PubMed: 22273582]
17. Olano C, Méndez C, Salas JA. Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Nat Prod Rep*. 2010; 27:571–616. [PubMed: 20352666]
18. Eustáquio AS, et al. Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from *S*-adenosyl-L-methionine. *Proc Natl Acad Sci USA*. 2009; 106:12295–12300. [PubMed: 19590008]
19. Mo S, et al. Biosynthesis of the allylmalonyl-CoA extender unit for the FK506 polyketide synthase proceeds through a dedicated polyketide synthase and facilitates the mutasynthesis of analogues. *J Am Chem Soc*. 2011; 133:976–985. [PubMed: 21175203]
20. Zhu X, Liu J, Zhang W. De novo biosynthesis of terminal alkyne-labeled natural products. *Nat Chem Biol*. 2015; 11:115–120. [PubMed: 25531891]
21. Yan Y, et al. Multiplexing of combinatorial chemistry in antimycin biosynthesis: expansion of molecular diversity and utility. *Angew Chem Int Ed Engl*. 2013; 52:12308–12312. [PubMed: 24123503]
22. Sundermann U, et al. Enzyme-directed mutasynthesis: a combined experimental and theoretical approach to substrate recognition of a polyketide synthase. *ACS Chem Biol*. 2013; 8:443–450. [PubMed: 23181268]
23. Koryakina I, et al. Poly specific *trans*-acyltransferase machinery revealed *via* engineered acyl-CoA synthetases. *ACS Chem Biol*. 2013; 8:200–208. [PubMed: 23083014]
24. Thirlway J, et al. Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. *Angew Chem Int Ed Engl*. 2012; 51:7181–7184. [PubMed: 22711659]
25. Kries H, et al. Reprogramming nonribosomal peptide synthetases for “clickable” amino acids. *Angew Chem Int Ed Engl*. 2014; 53:10105–10108. [PubMed: 25081643]
26. Evans BS, Chen Y, Metcalf WW, Zhao H, Kelleher NL. Directed evolution of the nonribosomal peptide synthetase AdmK generates new andrimid derivatives *in vivo*. *Chem Biol*. 2011; 18:601–607. [PubMed: 21609841]
27. Piel J. Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat Prod Rep*. 2010; 27:996–1047. [PubMed: 20464003]
28. Walker MC, et al. Expanding the fluorine chemistry of living systems using engineered polyketide synthase pathways. *Science*. 2013; 341:1089–1094. [PubMed: 24009388]
29. Dunn BJ, Khosla C. Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. *J R Soc Interface*. 2013; 10:20130297. [PubMed: 23720536]
30. Fischbach MA, Lai JR, Roche ED, Walsh CT, Liu DR. Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. *Proc Natl Acad Sci USA*. 2007; 104:11951–11956. [PubMed: 17620609]

31. Tanovic A, Samel SA, Essen LO, Marahiel MA. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science*. 2008; 321:659–663. [PubMed: 18583577]
32. Dutta S, et al. Structure of a modular polyketide synthase. *Nature*. 2014; 510:512–517. [PubMed: 24965652]
33. Whicher JR, et al. Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature*. 2014; 510:560–564. [PubMed: 24965656]
34. Johnson MN, Londergan CH, Charkoudian LK. Probing the phosphopantetheine arm conformations of acyl carrier proteins using vibrational spectroscopy. *J Am Chem Soc*. 2014; 136:11240–11243. [PubMed: 25080832]
35. Ross AC, et al. Biosynthetic multitasking facilitates thalassospiramide structural diversity in marine bacteria. *J Am Chem Soc*. 2013; 135:1155–1162. [PubMed: 23270364]
36. Kapur S, et al. Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc Natl Acad Sci USA*. 2012; 109:4110–4115. [PubMed: 22371562]
37. Gokhale RS, Tsuji SY, Cane DE, Khosla C. Dissecting and exploiting intermodular communication in polyketide synthases. *Science*. 1999; 284:482–485. [PubMed: 10205055]
38. Broadhurst RW, et al. The structure of docking domains in modular polyketide synthases. *Chem Biol*. 2003; 10:723–731. [PubMed: 12954331]
39. Hahn M, Stachelhaus T. Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proc Natl Acad Sci USA*. 2004; 101:15585–15590. [PubMed: 15498872]
40. Whicher JR, et al. Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem Biol*. 2013; 20:1340–1351. [PubMed: 24183970]
41. Ongley SE, Bian X, Neilan BA, Müller R. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat Prod Rep*. 2013; 30:1121–1138. [PubMed: 23832108]
42. Charlop-Powers Z, Milshteyn A, Brady SF. Metagenomic small molecule discovery methods. *Curr Opin Microbiol*. 2014; 19:70–75. [PubMed: 25000402]
43. Bachmann BO, Van Lanen SG, Baltz RH. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J Ind Microbiol Biotechnol*. 2014; 41:175–184. [PubMed: 24342967]
44. Thaker MN, Wright GD. Opportunities for synthetic biology in antibiotics: expanding glycopeptide chemical diversity. *ACS Synth Biol*. 2015; 4:195–206. [PubMed: 23654249]
45. Cobb RE, Ning JC, Zhao H. DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. *J Ind Microbiol Biotechnol*. 2014; 41:469–477. [PubMed: 24127070]
46. Gibson DG. Programming biological operating systems: genome design, assembly and activation. *Nat Methods*. 2014; 11:521–526. [PubMed: 24781325]
47. Zhang Y, Muyrers JP, Testa G, Stewart AF. DNA cloning by homologous recombination in *Escherichia coli*. *Nat Biotechnol*. 2000; 18:1314–1317. [PubMed: 11101815]
48. Fu J, et al. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat Biotechnol*. 2012; 30:440–446. [PubMed: 22544021]
49. Kouprina N, Larionov V. TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nat Rev Genet*. 2006; 7:805–812. [PubMed: 16983376]
50. Feng Z, Kim JH, Brady SF. Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type II PKS gene cluster. *J Am Chem Soc*. 2010; 132:11902–11903. [PubMed: 20690632]
51. Feng Z, Kallifidas D, Brady SF. Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *Proc Natl Acad Sci USA*. 2011; 108:12629–12634. [PubMed: 21768346]
52. Yamanaka K, et al. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc Natl Acad Sci USA*. 2014; 111:1957–1962. [PubMed: 24449899]

53. Gust B, Challis GL, Fowler K, Kieser T, Chater KF. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci USA*. 2003; 100:1541–1546. [PubMed: 12563033]
54. 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]
55. Gibson DG, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. 2009; 6:343–345. [PubMed: 19363495]
56. Li MZ, Elledge SJ. SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol Biol*. 2012; 852:51–59. [PubMed: 22328425]
57. Kushnir S, et al. Minimally invasive mutagenesis gives rise to a biosynthetic polyketide library. *Angew Chem Int Ed Engl*. 2012; 51:10664–10669. [PubMed: 22996590]
58. Song MC, et al. Microbial biosynthesis of medicinally important plant secondary metabolites. *Nat Prod Rep*. 2014; 31:1497–1509. [PubMed: 25072622]
59. Gomez-Escribano JP, Bibb MJ. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb Biotechnol*. 2011; 4:207–215. [PubMed: 21342466]
60. 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 USA*. 2010; 107:2646–2651. [PubMed: 20133795]
61. Zhou M, et al. Sequential deletion of all the polyketide synthase and nonribosomal peptide synthetase biosynthetic gene clusters and a 900-kb subtelomeric sequence of the linear chromosome of *Streptomyces coelicolor*. *FEMS Microbiol Lett*. 2012; 333:169–179. [PubMed: 22670631]
62. Komatsu M, et al. Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. *ACS Synth Biol*. 2013; 2:384–396. [PubMed: 23654282]
63. Rodriguez E, et al. Rapid engineering of polyketide overproduction by gene transfer to industrially optimized strains. *J Ind Microbiol Biotechnol*. 2003; 30:480–488. [PubMed: 12698320]
64. Li C, Hazzard C, Florova G, Reynolds KA. High titer production of tetracenomycins by heterologous expression of the pathway in a *Streptomyces cinnamonensis* industrial monensin producer strain. *Metab Eng*. 2009; 11:319–327. [PubMed: 19595787]
65. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science*. 2001; 291:1790–1792. [PubMed: 11230695]
66. Zhang H, Wang Y, Wu J, Skalina K, Pfeifer BA. Complete biosynthesis of erythromycin A and designed analogs using *E. coli* as a heterologous host. *Chem Biol*. 2010; 17:1232–1240. [PubMed: 21095573]
67. Watanabe K, et al. Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*. *Nat Chem Biol*. 2006; 2:423–428. [PubMed: 16799553]
68. Bian X, et al. Direct cloning, genetic engineering, and heterologous expression of the syringolin biosynthetic gene cluster in *E. coli* through Red/ET recombineering. *Chembiochem*. 2012; 13:1946–1952. [PubMed: 22851214]
69. Ross AC, Gulland LE, Dorrestein PC, Moore BS. Targeted capture and heterologous expression of the *Pseudoalteromonas* alterochromide gene cluster in *Escherichia coli* represents a promising natural product exploratory platform. *ACS Synth Biol*. 2015; 4:414–420. [PubMed: 25140825]
70. Liu J, Zhu X, Seipke RF, Zhang W. Biosynthesis of antimycins with a reconstituted 3-formamidosalicylate pharmacophore in *Escherichia coli*. *ACS Synth Biol*. 2015; 4:559–565. [PubMed: 25275920]
71. Ro DK, et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*. 2006; 440:940–943. [PubMed: 16612385]
72. Brown S, Clastre M, Courdavault V, O'Connor SE. De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc Natl Acad Sci USA*. 2015; 112:3205–3210. [PubMed: 25675512]
73. Ma SM, et al. Complete reconstitution of a highly reducing iterative polyketide synthase. *Science*. 2009; 326:589–592. [PubMed: 19900898]

74. Chiang YM, et al. An efficient system for heterologous expression of secondary metabolite genes in *Aspergillus nidulans*. *J Am Chem Soc.* 2013; 135:7720–7731. [PubMed: 23621425]
75. Mutka SC, Carney JR, Liu Y, Kennedy J. Heterologous production of epothilone C and D in *Escherichia coli*. *Biochemistry.* 2006; 45:1321–1330. [PubMed: 16430229]
76. Oßwald C, et al. Modular construction of a functional artificial epothilone polyketide pathway. *ACS Synth Biol.* 2014; 3:759–772. [PubMed: 23654254]
77. Dueber JE, et al. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol.* 2009; 27:753–759. [PubMed: 19648908]
78. Delebecque CJ, Lindner AB, Silver PA, Aldaye FA. Organization of intracellular reactions with rationally designed RNA assemblies. *Science.* 2011; 333:470–474. [PubMed: 21700839]
79. Conrado RJ, et al. DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic Acids Res.* 2012; 40:1879–1889. [PubMed: 22021385]
80. Temme K, et al. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res.* 2012; 40:8773–8781. [PubMed: 22743271]
81. Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol.* 2009; 27:946–950. [PubMed: 19801975]
82. Siegl T, Tokovenko B, Myronovskiy M, Luzhetskyy A. Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. *Metab Eng.* 2013; 19:98–106. [PubMed: 23876413]
83. Ran FA, et al. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc.* 2013; 8:2281–2308. [PubMed: 24157548]
84. Wang HH, et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature.* 2009; 460:894–898. [PubMed: 19633652]
85. Cong L, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science.* 2013; 339:819–823. [PubMed: 23287718]
86. Bao Z, et al. Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. *ACS Synth Biol.* 2014.10.1021/sb500255k
87. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol.* 2013; 31:233–239. [PubMed: 23360965]
88. Cobb RE, Wang Y, Zhao H. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth Biol.* 2014.10.1021/sb500351f
89. Uguru GC, et al. Synthetic RNA silencing of actinorhodin biosynthesis in *Streptomyces coelicolor* A3(2). *PLoS One.* 2013; 8:e67509. [PubMed: 23826310]
90. Na D, et al. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nat Biotechnol.* 2013; 31:170–174. [PubMed: 23334451]
91. Deb Roy A, Grüşchow S, Cairns N, Goss RJ. Gene expression enabling synthetic diversification of natural products: chemogenetic generation of pacidamycin analogs. *J Am Chem Soc.* 2010; 132:12243–12245. [PubMed: 20712319]
92. Arnison PG, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep.* 2013; 30:108–160. [PubMed: 23165928]
93. Park SR, Park JW, Ban YH, Sohng JK, Yoon YJ. 2-Deoxystreptamine-containing aminoglycoside antibiotics: recent advances in the characterization and manipulation of their biosynthetic pathways. *Nat Prod Rep.* 2013; 30:11–20. [PubMed: 23179168]
94. Kirby J, Keasling JD. Metabolic engineering of microorganisms for isoprenoid production. *Nat Prod Rep.* 2008; 25:656–661. [PubMed: 18663389]
95. Xu W, Gavia DJ, Tang Y. Biosynthesis of fungal indole alkaloids. *Nat Prod Rep.* 2014; 31:1474–1487. [PubMed: 25180619]
96. Schmidt EW, et al. Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. *Proc Natl Acad Sci USA.* 2005; 102:7315–7320. [PubMed: 15883371]
97. Ruffner DE, Schmidt EW, Heemstra JR. Assessing the combinatorial potential of the RiPP cyanobactin *tru* pathway. *ACS Synth Biol.* 2015; 4:482–492. [PubMed: 25140729]

98. Medema MH, van Raaphorst R, Takano E, Breitling R. Computational tools for the synthetic design of biochemical pathways. *Nat Rev Microbiol.* 2012; 10:191–202. [PubMed: 22266781]
99. Birmingham WR, et al. Bioretrosynthetic construction of a didanosine biosynthetic pathway. *Nat Chem Biol.* 2014; 10:392–399. [PubMed: 24657930]

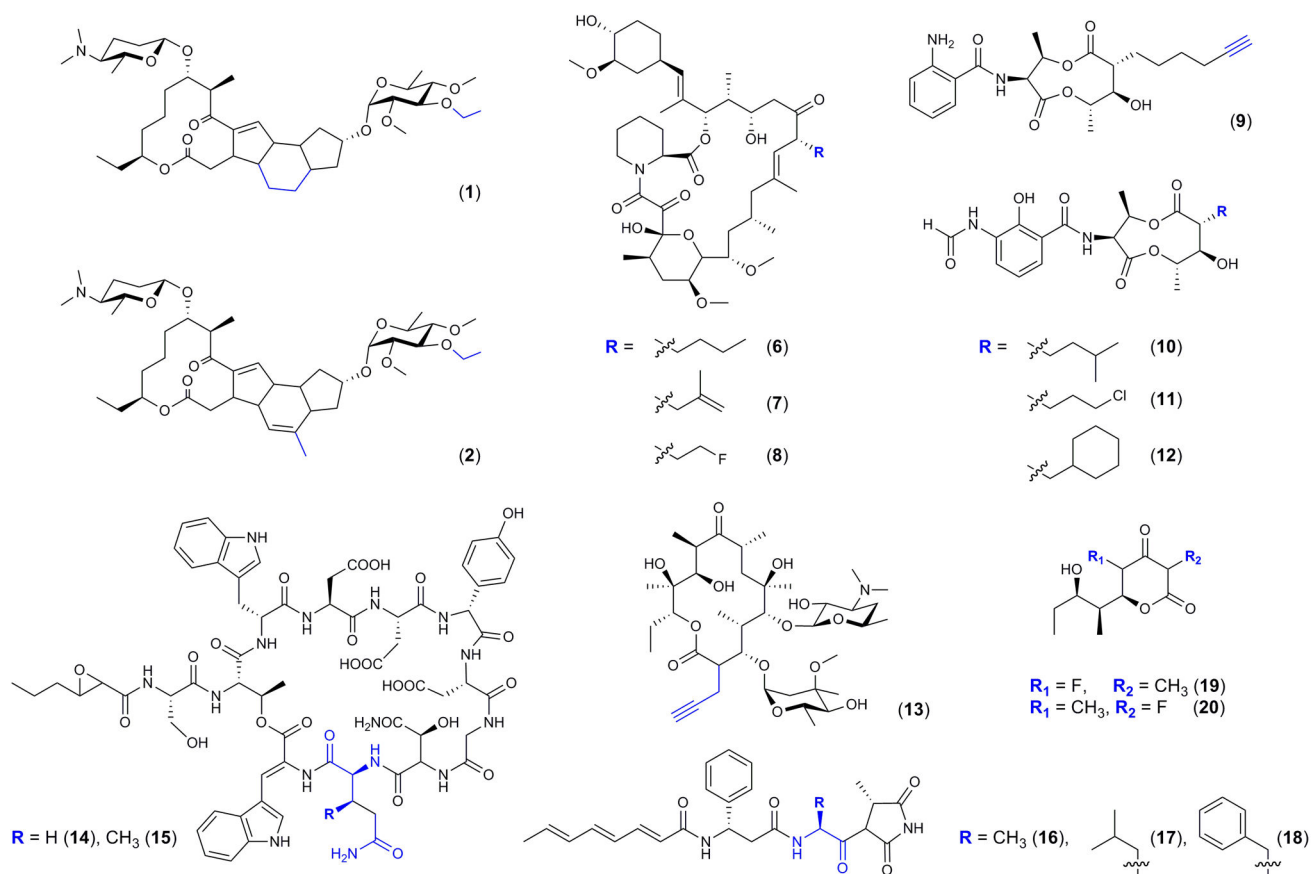


Figure 1. Structures of novel natural products generated by engineering the substrate specificity of biosynthetic enzymes. Modified structures from natural products are highlighted in blue.

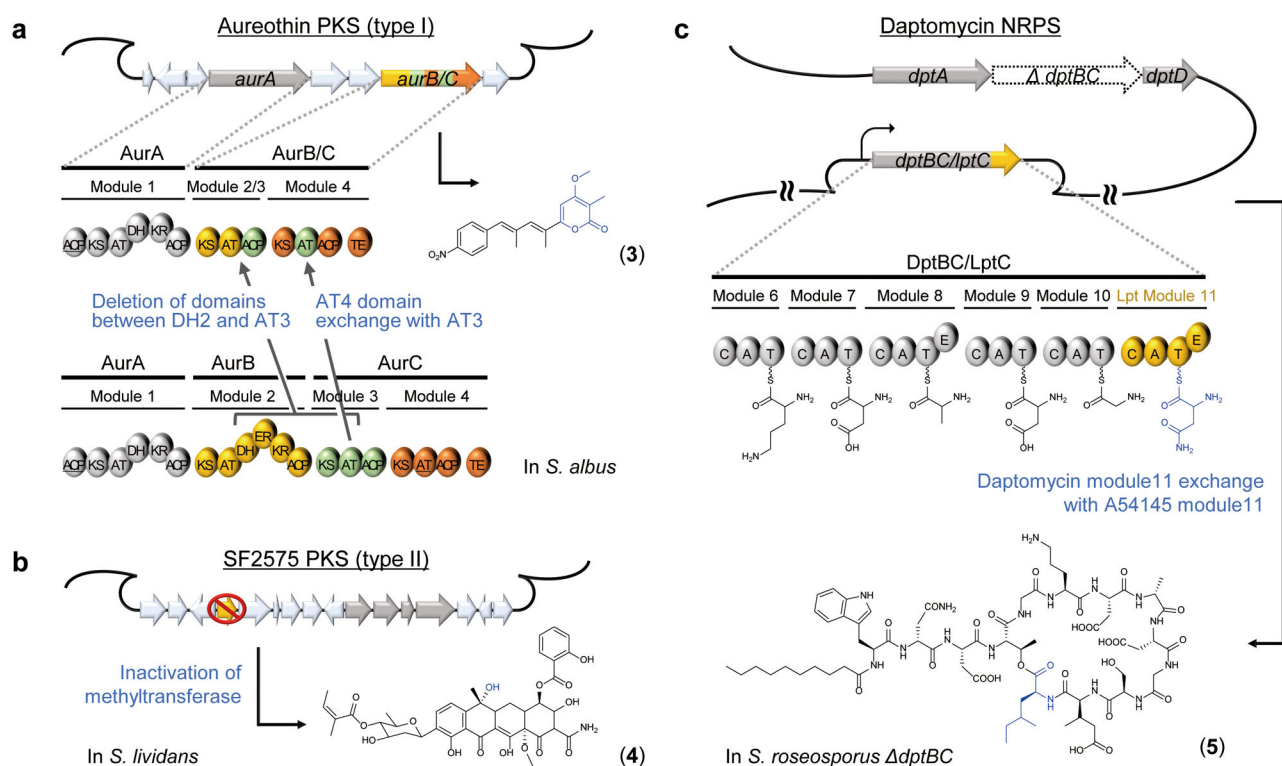


Figure 2. General strategies for the combinatorial biosynthesis of PKs and NRPs. **(a)** Through domain truncation and AT domain exchange, the type I aureotin PKS was morphed to produce its homologue luteoreticulin **(3)** in the heterologous host *S. albus*. **(b)** The tetracycline SF2575 biosynthetic pathway harboring a type II PKS and modifying enzymes was reconstituted in the heterologous host *S. lividans*. Subsequent inactivation of the methyltransferase gene generated tetracycline analogue **(4)**. **(c)** Novel lipopeptide **(5)** was produced by replacement of module 11 of daptomycin NRPS (DptBC) with the corresponding module from related A15454 NRPS (LptC) in the native daptomycin-producing *S. roseosporus*. Modified structures from natural products are highlighted in blue. AT, acyltransferase domain, ACP, acyl carrier protein; DH, dehydratase domain; ER, enoyl reductase domain, KS, ketosynthase domain; KR, ketoreductase domain; TE, thioesterase domain. A, adenylation domain; C, condensation domain; T, thiolation domain; E, epimerization domain.

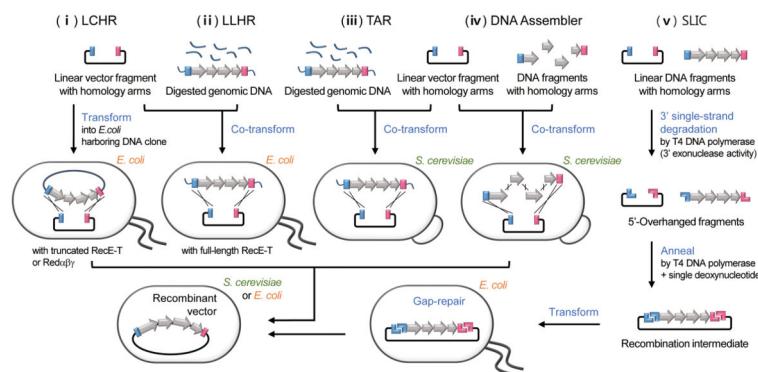
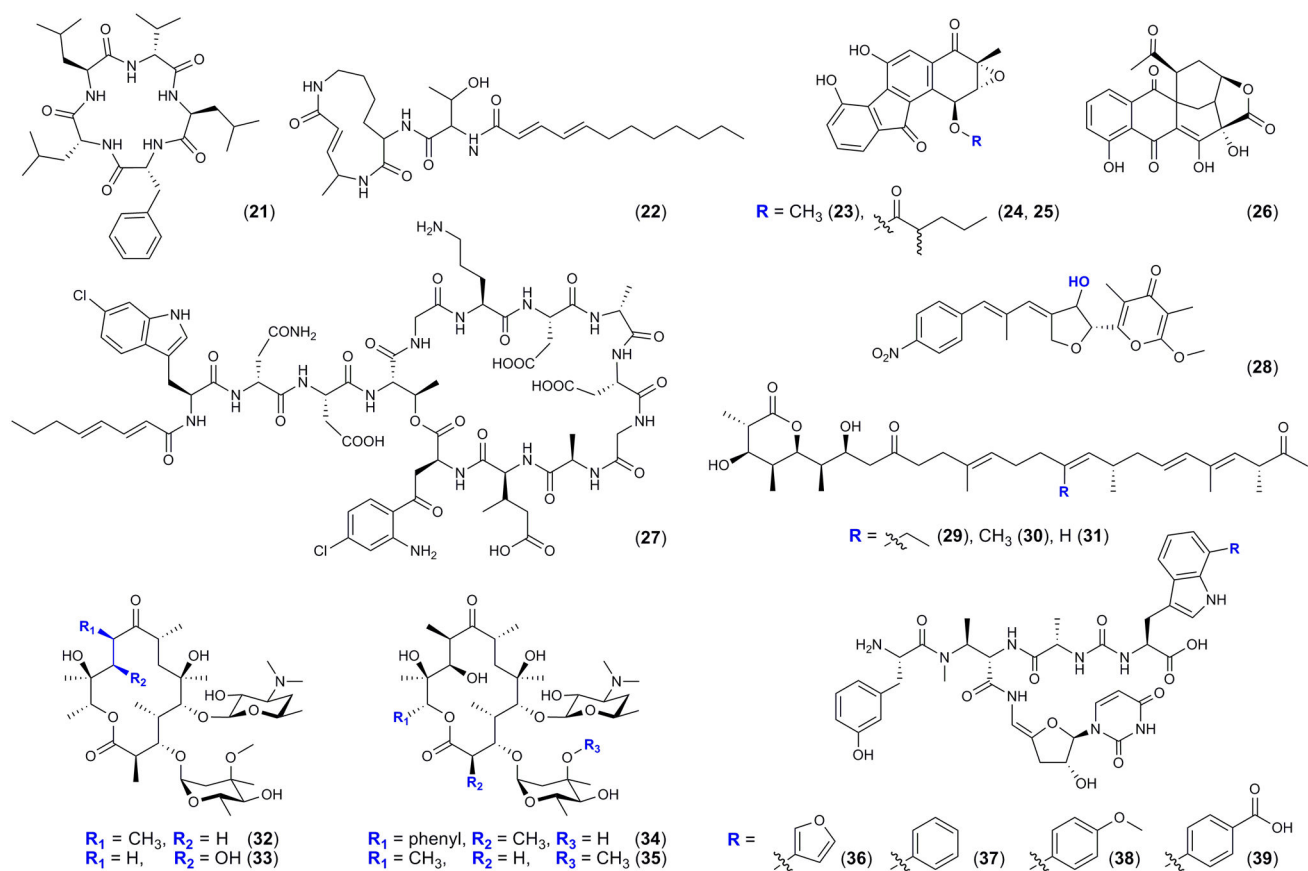


Figure 3. Approaches to assemble natural product biosynthetic pathways. See main text for details of each approach. LCHR, linear plus circular homologous recombination; LLHR, linear plus linear homologous recombination; SLIC, sequence and ligation independent cloning; TAR, transformation-associated recombination.

**Figure 4.**

The structures of novel or cryptic natural products produced by heterologous expression or chemobiosynthesis. Modified structures from natural products are highlighted in blue.

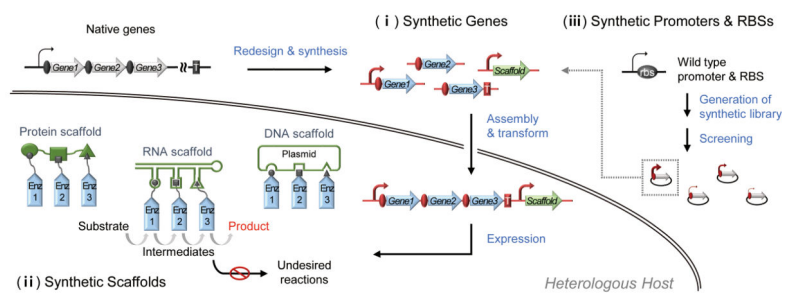


Figure 5. Representative synthetic biology tools for optimization of the expression of combinatorially assembled biosynthetic machineries. See main text for details of each approach. RBS, ribosome-binding site; T, terminator.

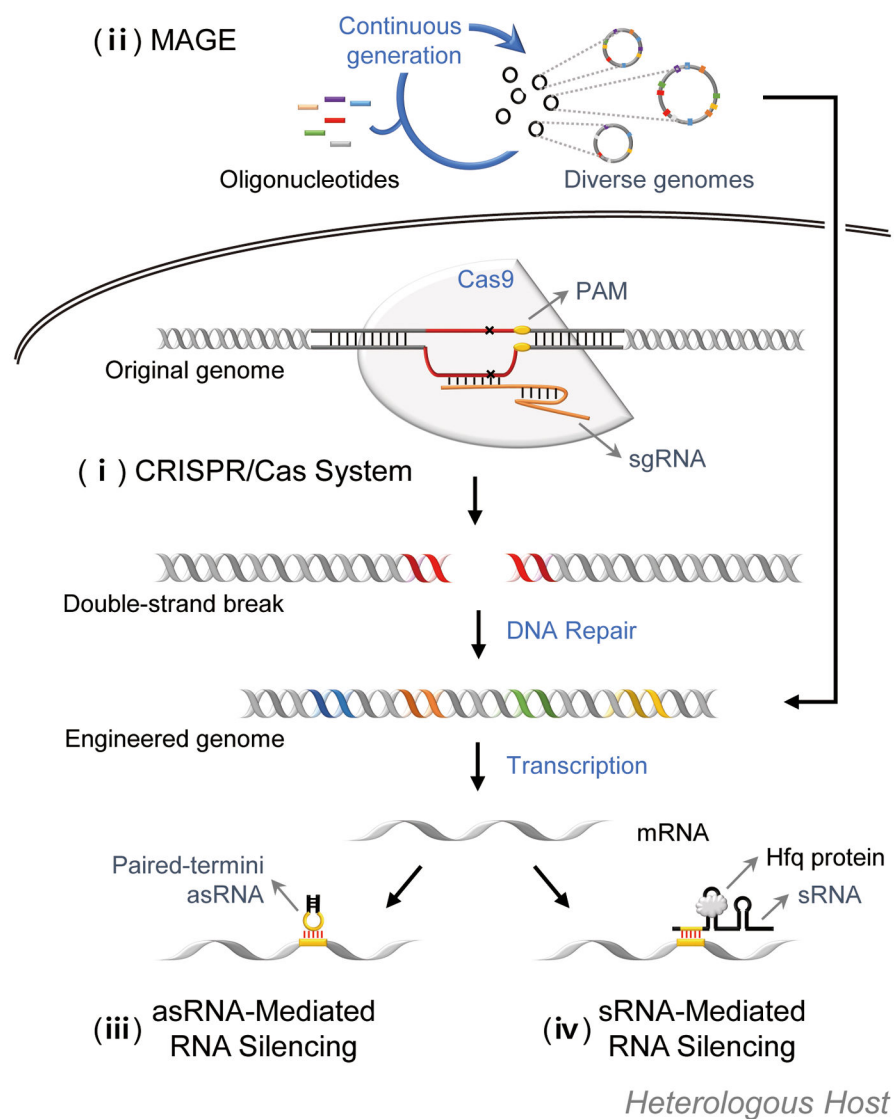


Figure 6. Representative synthetic biology tools for optimization of producing hosts. See main text for details of each approach. asRNA, antisense RNA; CRISPR/Cas, clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) protein; MAGE, multiplex automated genome engineering; PAM, protospacer-adjacent motif; sgRNA, single-guide RNA; sRNA, small regulatory RNA.