nature chemical biology

PERSPECTIVE

PUBLISHED ONLINE: 18 AUGUST 2015 | DOI: 10.1038/NCHEMBIO.1893

Reinvigorating natural product combinatorial biosynthesis with synthetic biology

Eunji Kim[1], Bradley S Moore[2,3]* & Yeo Joon Yoon[1]*

Natural products continue to play a pivotal role in drug-discovery efforts and in the understanding of human health. The ability to extend nature's chemistry through combinatorial biosynthesis—altering functional groups, regiochemistry and scaffold backbones through the manipulation of biosynthetic enzymes—offers unique opportunities to create natural product analogs. 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, the realities and the state of the art of combinatorial biosynthesis, including the engineering of substrate specificity of biosynthetic enzymes and the development of 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 had a fascination with controlling and extending nature's biosynthetic dexterity. The emergence in the 1980s of combinatorial chemistry 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 analogs of natural products differing by a single methyl group, by the oxidation state of a C-O bond, or by more drastic differences in 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 biosynthesis[1].

Motivation in the field was high, as researchers anticipated that numerous 'unnatural' natural products with altered structures could be produced, illuminating structure-activity relationships that are key for drug development purposes and improving pharmaceutical properties of clinically relevant compounds. Even today, the importance of natural products as drug leads cannot be overstated. They continue to account for the majority of antimicrobial and anticancer agents approved by the US Food and Drug Administration[2] while similarly serving as indispensable molecular probes in illuminating fundamental cellular processes[3]. Nevertheless, the pharmaceutical industry has severely cut back its support for natural products research in recent decades in favor of synthetic approaches to 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 libraries[4] and continue to cover a much wider chemical-space landscape.

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 spectrum[5], it is a mixture of 3′-*O*-ethyl-5,6-dihydrospinosyn J (**1**) and 3′-*O*-ethylspinosyn L (**2**) (**Fig. 1**) synthesized by the chemical modification of spinosyns J and L produced by a rhamnose 3′-*O*-methyltransferase–inactivated *Saccharopolyspora spinosa*[5]. Although spinetoram was not developed solely by combinatorial biosynthesis, this successful case strongly suggests that more potent natural product analogs can be discovered by combinatorial biosynthesis. Yet limited knowledge about biosynthetic enzymes and pathways, for instance of 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 could be programmed at will. We suggest that this gap is due to still-insufficient information about how individual biosynthetic enzymes function and how linked enzymatic modules function together and to limitations in the tools for rapidly testing relevant hypotheses. In this Perspective, we highlight successes in this area, identify open questions that are hindering further research and point to the 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 capabilities such as antibacterial, anticancer and immunosuppressant activities[6,7]. 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)[8], the aromatic PK actinorhodin from a type II PKS[9] and the lipopeptide antibiotic daptomycin from a NRP synthetase (NRPS)[10,11]. Hundreds of biosynthetic products (for example, compounds **3**–**5**; **Fig. 2**) 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[10–13] (**Fig. 2**). In the case of daptomycin, for instance, the combined use of these approaches has generated over 120 novel lipopeptides (such as **5**), some displaying improved therapeutic properties[10,11] (**Fig. 2c**).

[1]Department of Chemistry and Nano Science, Ewha Womans University, Seoul, Republic of Korea. [2]Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California, USA. [3]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, California, USA. *e-mail: bsmoore@ucsd.edu or joonyoon@ewha.ac.kr

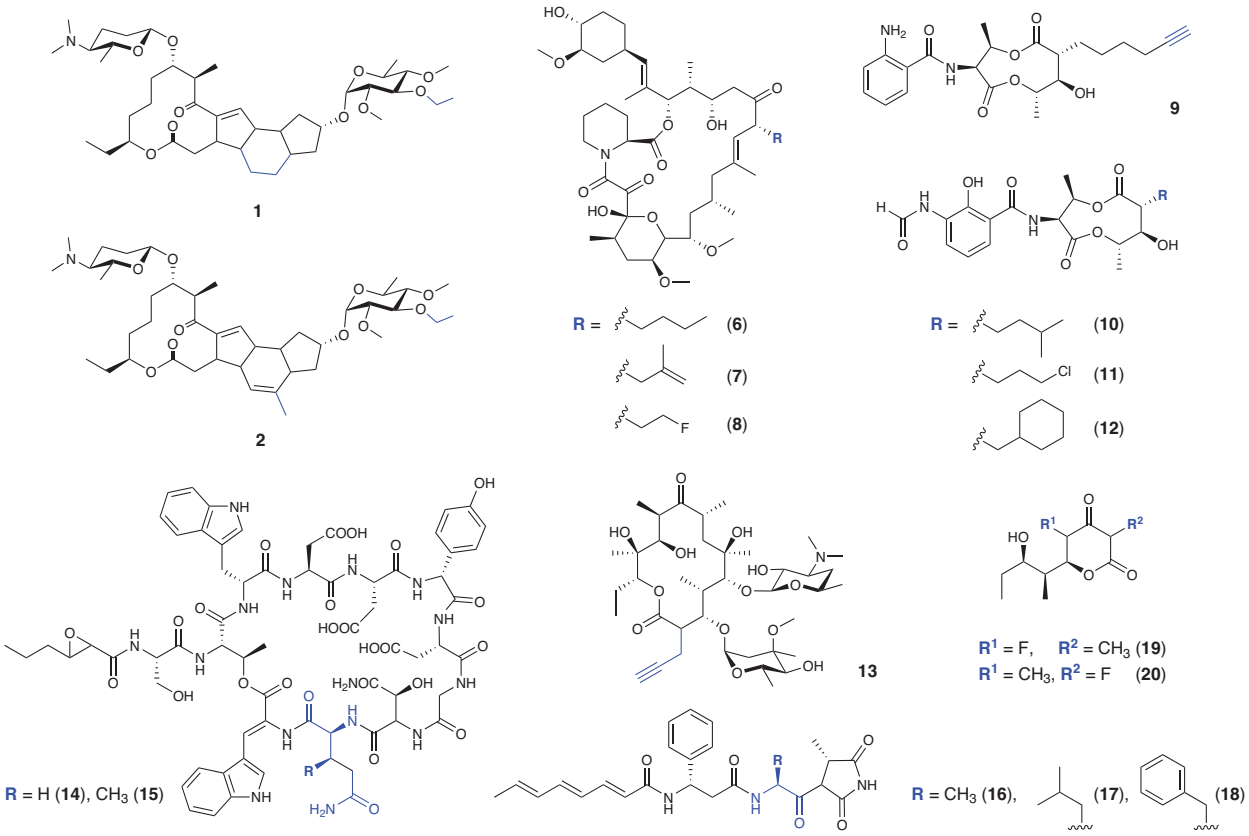

Figure 1 | Structures of novel natural products generated by engineering the substrate specificity of biosynthetic enzymes. Modified structures from natural products are highlighted in blue.

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 at 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 responsible 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 into PKSs[14] or nonproteinogenic amino acids into NRPSs[15]. Additional characterization of known types of tailoring enzymes that decorate these scaffolds, such as deoxyhexoses for modifying diverse biosynthetic aglycones[16], 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[15,16]. Unfortunately, the details of tailoring enzymes are beyond the scope of this article. Modification of 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. Over the past several years, however, 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 precursors[17]. This expansion to rare acyl-CoA extender units such as haloethylmalonyl-CoA, allylmalonyl-CoA, isobutyrylmalonyl-CoA and benzylmalonyl-CoA extends the structural possibilities for adding bulky or reactive side chains to the PK backbone[14]. 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**)[18]. Similarly, the AT domain in AntD of the antimycin NRPS-PKS hybrid assembly line showed remarkable tolerance toward a variety of non-natural extender units, including the engineered and 'clickable' 5-hexynylmalonyl-CoA[19] as well as linear, branched, cyclic and halogenated extender units[20]. This tolerance was exploited to produce diverse antimycin analogs (**9**–**12**) with 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 used to produce terminal alkyne-labeled natural products[19].

The incorporation of unnatural extender unit substrates, however, requires specific partner ATs to allow for their selective attachment to the modular assembly line. Although some ATs exhibit broad substrate scopes, most are specific and narrow in scope. 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. A 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**)[21] (**Fig. 1**). The expanded availability of new extender units derived from promiscuous CCRs[14] or malonyl-CoA synthetase (MatB)[22] with naturally occurring or engineered promiscuous AT domains provides great potential for the biosynthesis of novel PKs with specifically altered side chains.

In NRPS systems, the list of new nonproteinogenic amino acid substrates has ballooned to include hundreds of diverse structures. The modification of the native adenylation (A) domains that control the entry of diverse amino acid substrates into 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) and Glu to (2*S*,3*R*)-3-methyl-Gln (mGln) and Gln to produce the Gln- and mGln-containing CDA analogs (**14** and **15**) (**Fig. 1**)[23]. A similar point mutation in the Phe-specific A domain of the gramicidin synthetase GrsA (Trp239Ser) allowed the introduction of an *O*-propargyl-Tyr residue[24].

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 analogs 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*[25]. 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, the andrimid system is a relatively small NRPS, and the expandability of this structure-based screening method to multimodular NRPSs or PKSs needs to 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 analogs.

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 noninvasively engineered AT or A domains. An alternate approach

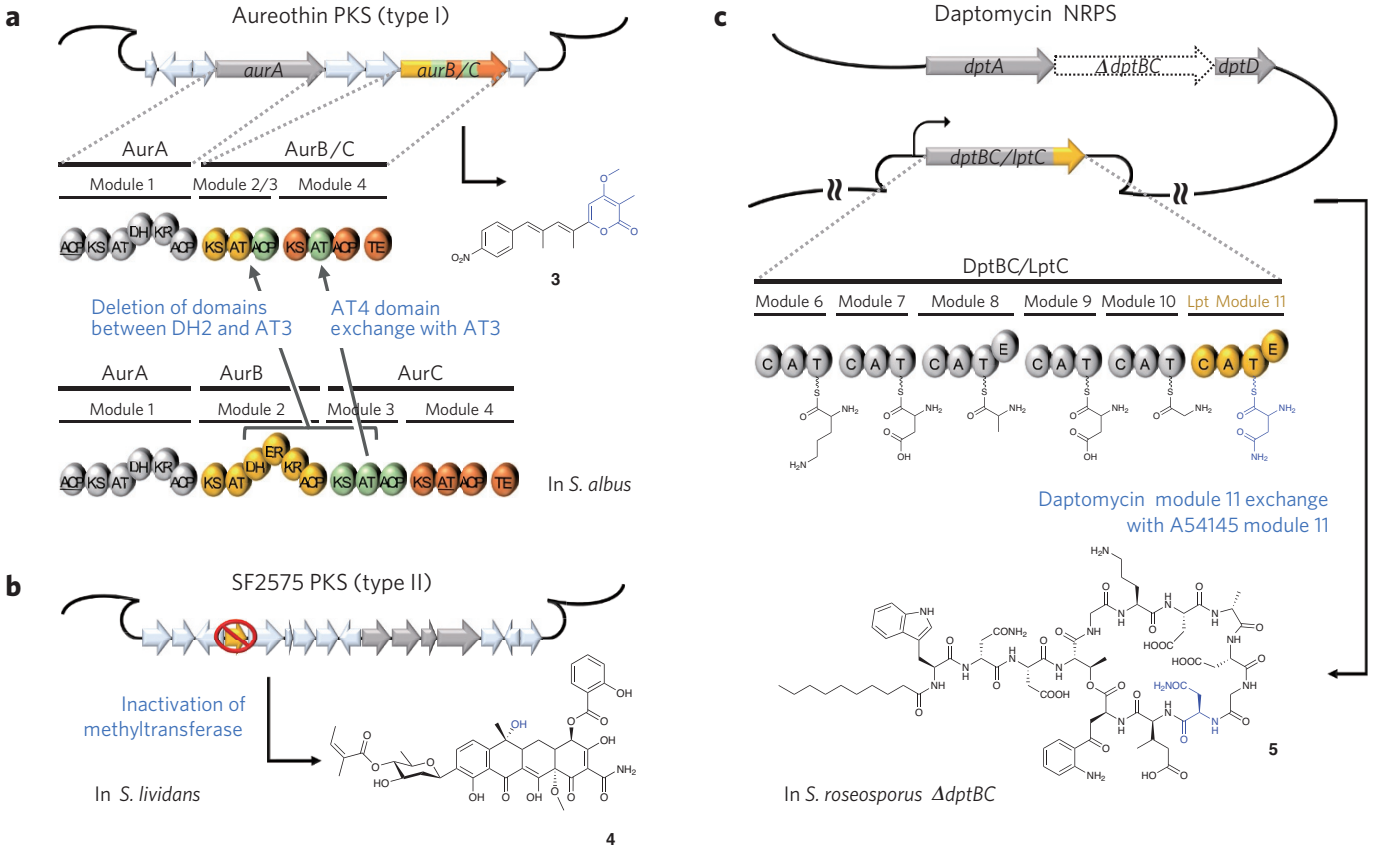

Figure 2 | General strategies for the combinatorial biosynthesis of PKs and NRPs. (**a**) Through domain truncation and AT domain exchange, the type I aureothin PKS was morphed to produce its homolog luteoreticulin (**3**) in the heterologous host *S. albus*. (**b**) The tetracycline SF2575 biosynthetic pathway harboring a type II PKS and modifying enzymes were reconstituted in the heterologous host *S. lividans*. Subsequent inactivation of the methyltransferase gene generated tetracycline analog (**4**). (**c**) Novel lipopeptide (**5**) was produced by replacement of module 11 of daptomycin NRPS (DptBC) with the corresponding module from related A54145 NRPS (LptC) in the native daptomycin-producing *Streptomyces 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.

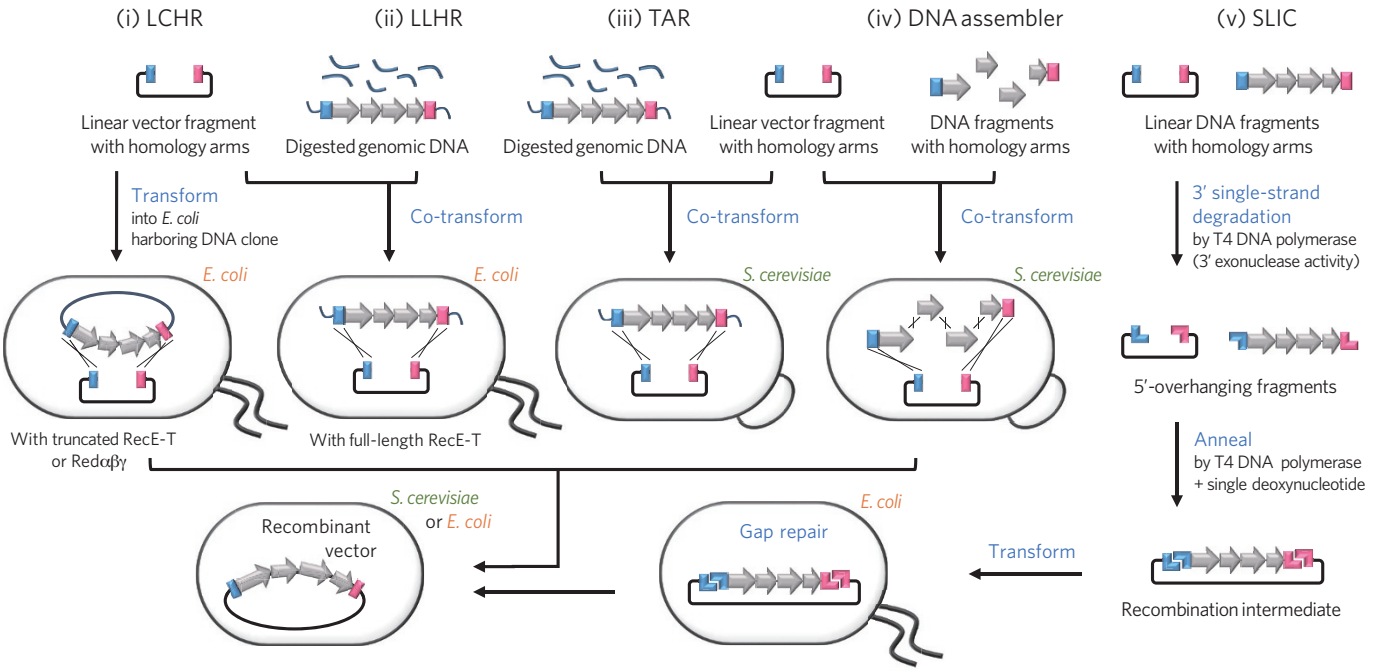

Figure 3 | Approaches to assembling natural product biosynthetic pathways. See main text for details of each approach. LCHR, linear-plus-circular homologous recombination; LLHR, linear-plus-linear homologous recombination; SLIC, sequence and ligation independent cloning; TAR, transformation-associated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

is to replace native domains with domains from other pathways that already contain the desired specificities. For example, orthogonal *trans*-AT PKSs, in which the AT domain is located on a freestanding enzyme instead of on the multimodular synthase[26], provide one successful case study of domain swapping. These domains are typically dedicated to malonyl-CoA, but the *trans*-AT protein KirCII from the kirromycin PKS is naturally specific for ethylmalonyl-CoA and possess a relaxed specificity toward non-natural extender units *in vitro*[22]. 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**)[27].

Although this example provides one success story for AT domain replacements, the strategy typically results in impaired protein folding or interactions that often significantly reduce biosynthetic productivity[28]. Efforts to create new NRP analogs by swapping A domains similarly result in reduced product yields as compared to those of native systems[10,11,15]. Although it has been demonstrated that directed evolution can improve the productivity of an engineered NRPS replaced with a heterologous A domain[29], 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: that is, the mode, timing, and consequences of the inter- and intramolecular interactions of megasynthases. Recent structural studies of NRPS[30] and PKS assembly lines[31,32] have provided insight into some of the intramolecular contacts and motions of these proteins (see p. 660). A recently developed thiocyanate vibrational spectroscopic probe installed on the terminal thiol of the ACP's 4′-phosphopantetheine arm[33], which enables the observation of ACP conformational changes during biosynthesis, will also be a useful tool for comprehending PKS operating principles and designing 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 in which each and every enzymatic domain is used once per assembled product[6,7]. In some cases, however, enzymatic domains or modules are entirely skipped or are 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 reactions[34]. 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 PKS[35]. Based on this model, module 3 of the erythromycin PKS was reprogramed 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 3 (ref. 35). Additionally, the corresponding PKS and NRPS subunits must interact with their appropriate partners, and these interactions are mediated in part by docking domains[36,37] and communication-mediating domains[38], 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 mutagenesis[39]. 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, a major impediment 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[40–42]. Multidisciplinary synthetic biology approaches are poised to enable combinatorial biosynthesis studies to become faster and more accurate[15,43] 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[44,45]. Homology-based methods have been extensively used for the heterologous expression of natural product gene clusters, and studies in which 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) that is catalyzed by Redαβγ from the coliphage λ Red system or by a truncated version of RecE along with RecT from the Rac prophage (**Fig. 3**)[46]. LCHR was successfully used to exchange single or multiple domains for the combinatorial biosynthesis of the daptomycin derivatives described above[10,11]. Over the past decade, a range of natural product gene clusters have been heterologously expressed by this method[40]. In spite of this success, Red/ET recombineering becomes difficult when assembling large clusters cloned on multiple overlapping clones since stepwise recombination with a unique selectable maker 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 than LCHR (**Fig. 3**)[47]. LLHR was successfully used for the direct cloning of ten unknown PKS/NRPS gene clusters from *Photorhabdus luminescens* into *E. coli*, circumventing library construction, stepwise 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**)[47]. Although LLHR allows 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*[48] have been developed to directly capture gene clusters from environmental DNA samples or large genomic loci (**Fig. 3**). By using 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**) (ref. 49) and pentacyclic ring **26** (ref. 50) 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* because it contains the cosmid-derived pUC origin of replication. 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 *Streptomyces coelicolor*, yielding

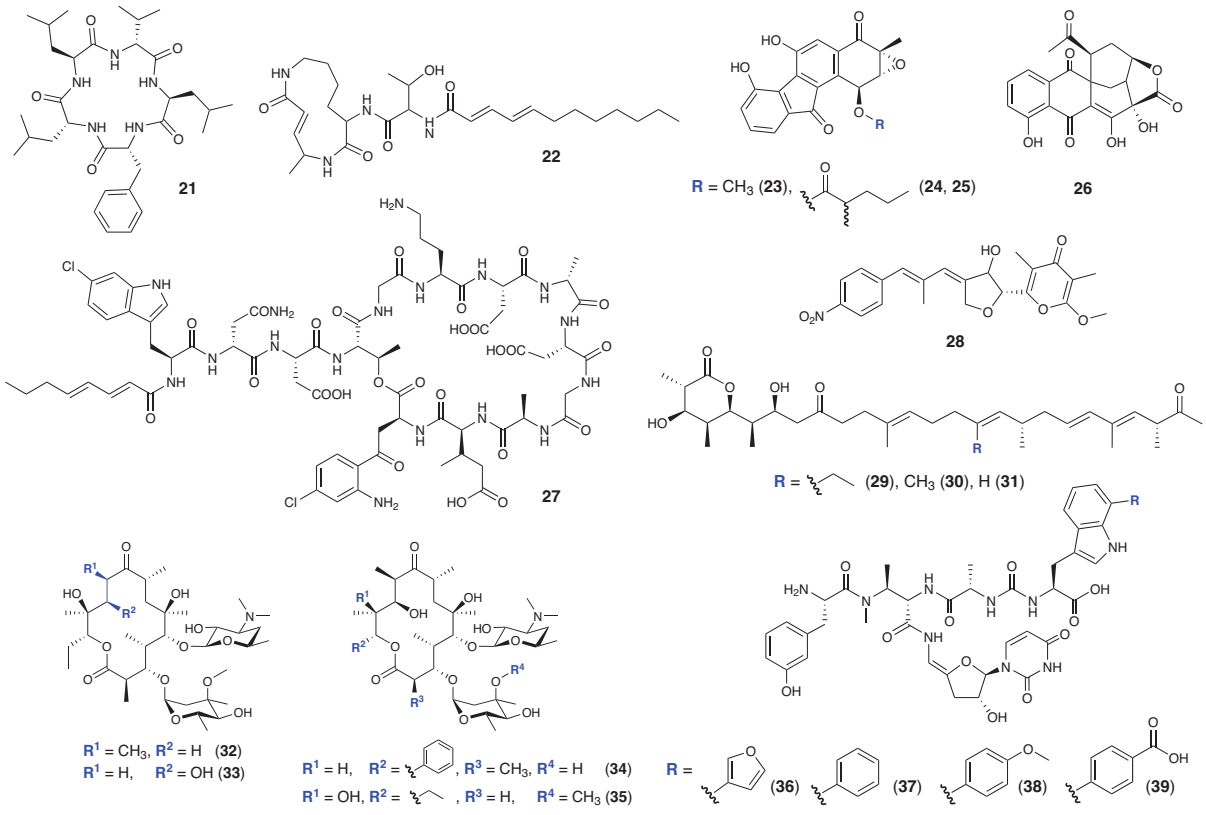

Figure 4 | The structures of novel or cryptic natural products produced by heterologous expression or chemobiosynthesis. Modified structures from natural products are highlighted in blue.

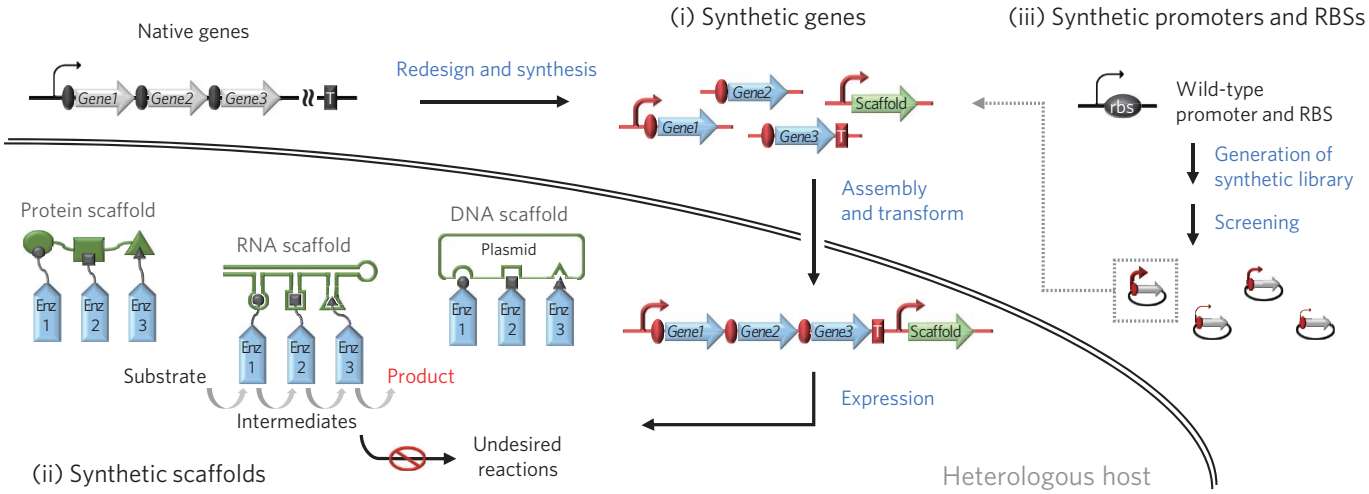

Figure 5 | Representative synthetic biology tools for optimization of the expression of combinatorially assembled biosynthetic machineries. See main text for details of each approach. RBS, ribosome-binding site; T, terminator.

a dihalogenated derivative of daptomycin, taromycin A (**27**) (ref. 51) (**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 targeting[52] 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- to 5-kb fragments using the DNA assembler method and was expressed in *Streptomyces lividans*, producing aureothin[53]. In addition, because 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 the aureothin analog **28** (ref. 53) (**Fig. 4**). However, with the DNA assembler method, random mutations are likely to occur as a result of 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"[54], 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; these 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)[55] is one variation of this *in vitro* homology-based method, in which T4 DNA polymerase is used for both 3′ single-strand degradation and partial gap-filling (**Fig. 3**). SLIC has been applied to the combinatorial biosynthesis of premonensin derivatives[56]. Mutated fragments of β-carbon processing domains of the 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, yielding a library of 22 premonensin redox derivatives of which a few (**29**–**31**) (**Fig. 4**) showed increased antibacterial activity against *Pseudomonas aeruginosa*[56]. Although the construct size that can be assembled *in vitro* seem to be smaller than that *in vivo*, a distinct advantage of *in vitro* DNA assembly is that it can be accomplished in hours, whereas *in vivo* methods can take days.

Toward optimal chassis for heterologous biosynthesis

Sequences from diverse bacteria, such as myxobacteria and cyanobacteria[40], from fungi[40] and from plants[57], and even environmental DNA[41], 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 publicly 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 the 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* (for example, having high GC content and Gram positive), 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 (for example, phosphopantetheinyl transferase (PPTase) responsible for adding a 4′-phosphopantetheine group to PKS and NRPS carrier proteins) that is required for functional biosynthetic enzymes. Although 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 the most commonly employed. Removal of endogenous secondary metabolite genes from the chassis strain is advantageous because precursors and energy are then redirected for the production of the heterologous compound and because a cleaner metabolite background simplifies the detection and purification of the heterologous compound. To this end, variant strains of *S. coelicolor*[58] and *S. avermitilis*[59] 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[58–61].

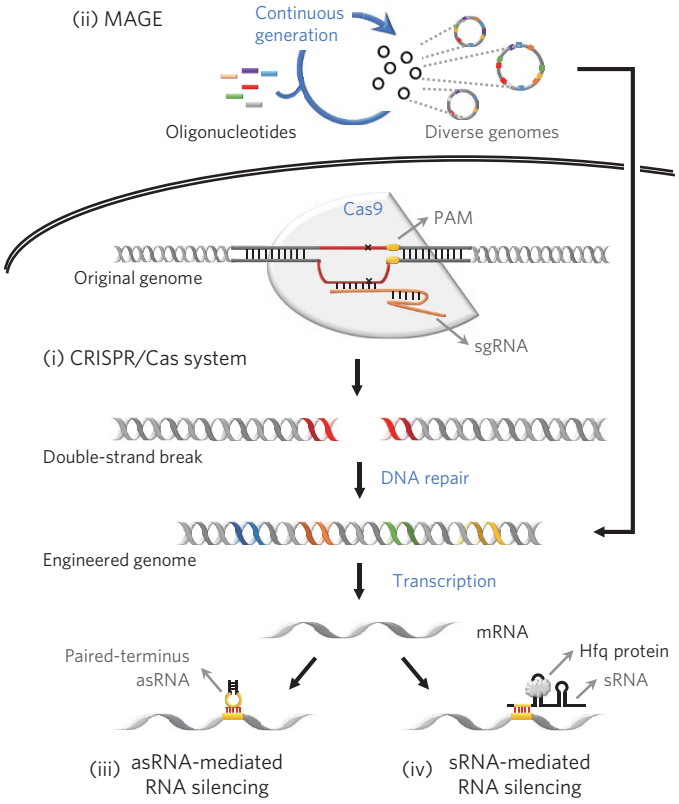

Figure 6 | Representative synthetic biology tools for optimization of producing hosts. See main text for details of each approach. asRNA, antisense RNA; CRISPR/Cas, clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) protein; MAGE, multiplex automated genome engineering;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

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 the *Saccharopolyspora erythraea* industrial strain were exploited to enhance titers of heterologous PKs. An approximately ten-fold higher production of erythromycin analogs (**32** and **33**) (**Fig. 4**) was observed in the engineered *S. erythraea* overproducer by chromosomal integration of engineered PKSs as compared to *S. lividans*[62]. 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 PKSs. Multi-gram quantities of tetracenomycins were produced by the heterologous expression of type II tetracenomycin PKS genes from *Streptomyces glaucescens* in an *S. cinnamonensis* strain producing high levels of monensin A[63]. 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. Nonetheless, 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 post-translationally modifying PKS ACP domains and of supplying methylmalonyl-CoA in sufficient amounts[64]. 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 resistance[65]. This *E. coli* platform was further applied to the production of erythromycin analogs (**34** and **35**) (**Fig. 4**) by replacing the loading module and altering the extender unit specificity of the second module, respectively[65]. NRPs such as echinomycin from *Streptomyces lasaliensis*[66], syringolin derivatives derived from *Pseudomonas syringae*[67] and alterochromide lipopeptides from *Pseudoalteromonas piscicida* JCM 20779 (ref. 68), as well as the NRP-PK hybrid antimycin from *S. albus* or *Streptomyces ambofaciens*[69], were also produced in *E. coli*. These results demonstrate that *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 products[40]. 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 proved to be a promising approach to access plant metabolites such as the terpenoid antimalarial drug precursor artemisinic acid[70] and strictosidine, the last common biosynthetic intermediate for all monoterpene indole alkaloids[71]. The iterative fungal PKS LovB from the lovastatin pathway of *Aspergillus terreus* was also successfully reconstituted in *S. cerevisiae*[72]. 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 asperfuranone[73]. 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 to be capable of supplying the precursors for the targeted secondary metabolites and efficiently expressing the target biosynthetic pathways in a controllable manner (for example, *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 existing synthetic biology tools are designed for *E. coli* and may not be directly applicable to combinatorial biosynthesis in actinomycetes, advancements in synthetic biology methodology offer new approaches to address the problems associated with combinatorial biosynthesis (**Figs. 5** and **6**). For example, it is now possible to use synthetic DNA to functionally reconstitute large PKS-NRPS pathways, thereby enabling the use of optimized codons in desirable hosts (**Fig. 5**). A 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 so as 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* BAP1 (ref. 74). 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*[75]. Although the reconstitution of large PKS or 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 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*[76]. 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 that of a system without the scaffold[77]. 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 five-fold) of resveratrol, 1,2-propanediol and mevalonate in *E. coli*[78]. 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 promoters[79] and ribosome binding sites (RBSs)[80] 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 of *S. erythraea*, and an approximately three-fold increase in flaviolin production was observed as compared with that for the ermEp 1* promoter (a strong variant of ermEp 1) in *S. lividans* (**Fig. 5**)[81].

High-throughput genome editing tools, such as the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) protein (CRISPR/Cas) system[82], and genomic evolution technology, such as the multiplex automated genome engineering (MAGE)[83], 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) that 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 cells[84], *S. cerevisiae*[85], *E. coli*[86] and different *Streptomyces* species[87], 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. As an alternative, 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 the isolation of an *E. coli* variant with a more than five-fold increase in lycopene production within 3 days (**Fig. 6**)[83]. However, because this technology requires an efficient transformation and single-stranded oligonucleotide-mediated allelic replacement, it may not be amenable to 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**)[88]. 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, and can thus be used to control 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 tyrosine[89]. This method is useful for large-scale target identification and for modulating chromosomal gene expression without modifying these genes, but unfortunately, its applicability to hosts other than *E. coli* has not yet been shown.

When developed for use in a wider range of chassis, including the most prolific natural product–producing actinomycetes, the synthetic biology tools listed above will facilitate essential combinatorial biosynthesis studies aimed at broadening the scope of biosynthetic building blocks and 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 derivatives[4]. 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 (for example, **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 chemistry[20]. 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 analogs (**36**–**39**) (**Fig. 4**) using mild cross-coupling conditions in crude aqueous extracts of the culture broth[90]. 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 cooperate more closely to design useful analogs as well as to generate compounds that are difficult to make by solely chemical means.

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 RIPPs[91], saccharides[92], terpenoids[93] and alkaloids[94] from assorted organisms. Because the non-modular biosynthetic machineries for these natural products are more amenable to 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 patellamide[95] and cyanobactin pathways[96] 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 yeast[70].

With the increasing scientific understanding and technical capabilities gained over the last two decades, we argue that the basic foundation has been 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 retrosynthetic design of artificial biosynthetic genes using computational tools[97] (see p. 639) and allow them to be 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 analog 2′,3′-dideoxyinosine (didanosine) in a retrograde fashion[98]. 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 owing 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*[98].

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

Received 15 June 2015; accepted 22 July 2015; published online 18 August 2015; corrected after print 23 September 2015

References

1. Hopwood, D.A. *et al.* Production of 'hybrid' antibiotics by genetic engineering. *Nature* **314**, 642–644 (1985).
2. Newman, D.J. & Cragg, G.M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **75**, 311–335 (2012).
3. Robles, O. & Romo, D. Chemo- and site-selective derivatizations of natural products enabling biological studies. *Nat. Prod. Rep.* **31**, 318–334 (2014).
4. Li, J.W. & Vederas, J.C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **325**, 161–165 (2009).
5. Huang, K.X., Xia, L., Zhang, Y., Ding, X. & Zahn, J.A. Recent advances in the biochemistry of spinosyns. *Appl. Microbiol. Biotechnol.* **82**, 13–23 (2009).
6. Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Edn Engl.* **48**, 4688–4716 (2009).
7. Schwarzer, D., Finking, R. & Marahiel, M.A. Nonribosomal peptides: from genes to products. *Nat. Prod. Rep.* **20**, 275–287 (2003).
8. 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* **96**, 1846–1851 (1999).
9. McDaniel, R., Ebert-Khosla, S., Hopwood, D.A. & Khosla, C. Engineered biosynthesis of novel polyketides. *Science* **262**, 1546–1550 (1993).
10. Nguyen, K.T. *et al.* Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc. Natl. Acad. Sci. USA* **103**, 17462–17467 (2006).
11. Baltz, R.H. Combinatorial biosynthesis of cyclic lipopeptide antibiotics: a model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways. *ACS Synth. Biol.* **3**, 748–758 (2014).
12. Sugimoto, Y., Ding, L., Ishida, K. & Hertweck, C. Rational design of modular polyketide synthases: morphing the aureothin pathway into a luteoreticulin assembly line. *Angew. Chem. Int. Edn Engl.* **53**, 1560–1564 (2014).
13. Wang, P., Kim, W., Pickens, L.B., Gao, X. & Tang, Y. Heterologous expression and manipulation of three tetracycline biosynthetic pathways. *Angew. Chem. Int. Edn Engl.* **51**, 11136–11140 (2012).
14. Wilson, M.C. & Moore, B.S. Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Nat. Prod. Rep.* **29**, 72–86 (2012).
15. Giessen, T.W. & Marahiel, M.A. Ribosome-independent biosynthesis of biologically active peptides: Application of synthetic biology to generate structural diversity. *FEBS Lett.* **586**, 2065–2075 (2012).
16. Olano, C., Méndez, C. & Salas, J.A. Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Nat. Prod. Rep.* **27**, 571–616 (2010).
17. Eustáquio, A.S. *et al.* Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from *S*-adenosyl-L-methionine. *Proc. Natl. Acad. Sci. USA* **106**, 12295–12300 (2009).
18. 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.* **133**, 976–985 (2011).
19. Zhu, X., Liu, J. & Zhang, W. *De novo* biosynthesis of terminal alkyne–labeled natural products. *Nat. Chem. Biol.* **11**, 115–120 (2015).
20. Yan, Y. *et al.* Multiplexing of combinatorial chemistry in antimycin biosynthesis: expansion of molecular diversity and utility. *Angew. Chem. Int. Edn Engl.* **52**, 12308–12312 (2013).
21. Sundermann, U. *et al.* Enzyme-directed mutasynthesis: a combined experimental and theoretical approach to substrate recognition of a polyketide synthase. *ACS Chem. Biol.* **8**, 443–450 (2013).
22. Koryakina, I. *et al.* Poly specific *trans*-acyltransferase machinery revealed *via* engineered acyl-CoA synthetases. *ACS Chem. Biol.* **8**, 200–208 (2013).

23. 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. Edn Engl.* **51**, 7181–7184 (2012).
24. Kries, H. *et al.* Reprogramming nonribosomal peptide synthetases for "clickable" amino acids. *Angew. Chem. Int. Edn Engl.* **53**, 10105–10108 (2014).
25. Evans, B.S., Chen, Y., Metcalf, W.W., Zhao, H. & Kelleher, N.L. Directed evolution of the nonribosomal peptide synthetase AdmK generates new andrimid derivatives *in vivo*. *Chem. Biol.* **18**, 601–607 (2011).
26. Piel, J. Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat. Prod. Rep.* **27**, 996–1047 (2010).
27. Walker, M.C. *et al.* Expanding the fluorine chemistry of living systems using engineered polyketide synthase pathways. *Science* **341**, 1089–1094 (2013).
28. Dunn, B.J. & Khosla, C. Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. *J. R. Soc. Interface* **10**, 20130297 (2013).
29. Fischbach, M.A., Lai, J.R., Roche, E.D., Walsh, C.T. & Liu, D.R. Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. *Proc. Natl. Acad. Sci. USA* **104**, 11951–11956 (2007).
30. Tanovic, A., Samel, S.A., Essen, L.O. & Marahiel, M.A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659–663 (2008).
31. Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510**, 512–517 (2014).
32. Whicher, J.R. *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560–564 (2014).
33. Johnson, M.N., Londergan, C.H. & Charkoudian, L.K. Probing the phosphopantetheine arm conformations of acyl carrier proteins using vibrational spectroscopy. *J. Am. Chem. Soc.* **136**, 11240–11243 (2014).
34. Ross, A.C. *et al.* Biosynthetic multitasking facilitates thalassospiramide structural diversity in marine bacteria. *J. Am. Chem. Soc.* **135**, 1155–1162 (2013).
35. Kapur, S. *et al.* Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc. Natl. Acad. Sci. USA* **109**, 4110–4115 (2012).
36. Gokhale, R.S., Tsuji, S.Y., Cane, D.E. & Khosla, C. Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**, 482–485 (1999).
37. Broadhurst, R.W. *et al.* The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**, 723–731 (2003).
38. Hahn, M. & Stachelhaus, T. Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proc. Natl. Acad. Sci. USA* **101**, 15585–15590 (2004).
39. Whicher, J.R. *et al.* Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem. Biol.* **20**, 1340–1351 (2013).
40. Ongley, S.E., Bian, X., Neilan, B.A. & Müller, R. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep.* **30**, 1121–1138 (2013).
41. Charlop-Powers, Z., Milshteyn, A. & Brady, S.F. Metagenomic small molecule discovery methods. *Curr. Opin. Microbiol.* **19**, 70–75 (2014).
42. Bachmann, B.O., Van Lanen, S.G. & Baltz, R.H. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J. Ind. Microbiol. Biotechnol.* **41**, 175–184 (2014).
43. Thaker, M.N. & Wright, G.D. Opportunities for synthetic biology in antibiotics: expanding glycopeptide chemical diversity. *ACS Synth. Biol.* **4**, 195–206 (2015).
44. Cobb, R.E., Ning, J.C. & Zhao, H. DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. *J. Ind. Microbiol. Biotechnol.* **41**, 469–477 (2014).
45. Gibson, D.G. Programming biological operating systems: genome design, assembly and activation. *Nat. Methods* **11**, 521–526 (2014).
46. Zhang, Y., Muyrers, J.P., Testa, G. & Stewart, A.F. DNA cloning by homologous recombination in *Escherichia coli*. *Nat. Biotechnol.* **18**, 1314–1317 (2000).
47. Fu, J. *et al.* Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **30**, 440–446 (2012).
48. Kouprina, N. & Larionov, V. TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nat. Rev. Genet.* **7**, 805–812 (2006).
49. Feng, Z., Kim, J.H. & Brady, S.F. Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type II PKS gene cluster. *J. Am. Chem. Soc.* **132**, 11902–11903 (2010).
50. Feng, Z., Kallifidas, D. & Brady, S.F. Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *Proc. Natl. Acad. Sci. USA* **108**, 12629–12634 (2011).
51. 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* **111**, 1957–1962 (2014).
52. Gust, B., Challis, G.L., Fowler, K., Kieser, T. & Chater, K.F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. USA* **100**, 1541–1546 (2003).
53. Shao, Z., Luo, Y. & Zhao, H. Rapid characterization and engineering of natural product biosynthetic pathways *via* DNA assembler. *Mol. Biosyst.* **7**, 1056–1059 (2011).
54. Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
55. Li, M.Z. & Elledge, S.J. SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol. Biol.* **852**, 51–59 (2012).
56. Kushnir, S. *et al.* Minimally invasive mutagenesis gives rise to a biosynthetic polyketide library. *Angew. Chem. Int. Edn Engl.* **51**, 10664–10669 (2012).
57. Song, M.C. *et al.* Microbial biosynthesis of medicinally important plant secondary metabolites. *Nat. Prod. Rep.* **31**, 1497–1509 (2014).
58. Gomez-Escribano, J.P. & Bibb, M.J. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **4**, 207–215 (2011).
59. Komatsu, M., Uchiyama, T., Omura, S., Cane, D.E. & Ikeda, H. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. USA* **107**, 2646–2651 (2010).
60. 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.* **333**, 169–179 (2012).
61. Komatsu, M. *et al.* Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. *ACS Synth. Biol.* **2**, 384–396 (2013).
62. Rodriguez, E. *et al.* Rapid engineering of polyketide overproduction by gene transfer to industrially optimized strains. *J. Ind. Microbiol. Biotechnol.* **30**, 480–488 (2003).
63. Li, C., Hazzard, C., Florova, G. & Reynolds, K.A. High titer production of tetracenomycins by heterologous expression of the pathway in a *Streptomyces cinnamonensis* industrial monensin producer strain. *Metab. Eng.* **11**, 319–327 (2009).
64. Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790–1792 (2001).
65. Zhang, H., Wang, Y., Wu, J., Skalina, K. & Pfeifer, B.A. Complete biosynthesis of erythromycin A and designed analogs using *E. coli* as a heterologous host. *Chem. Biol.* **17**, 1232–1240 (2010).
66. Watanabe, K. *et al.* Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*. *Nat. Chem. Biol.* **2**, 423–428 (2006).
67. 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* **13**, 1946–1952 (2012).
68. Ross, A.C., Gulland, L.E., Dorrestein, P.C. & Moore, B.S. Targeted capture and heterologous expression of the *Pseudoalteromonas* alterochromide gene cluster in *Escherichia coli* represents a promising natural product exploratory platform. *ACS Synth. Biol.* **4**, 414–420 (2015).
69. Liu, J., Zhu, X., Seipke, R.F. & Zhang, W. Biosynthesis of antimycins with a reconstituted 3-formamidosalicylate pharmacophore in *Escherichia coli*. *ACS Synth. Biol.* **4**, 559–565 (2015).
70. Ro, D.K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
71. Brown, S., Clastre, M., Courdavault, V. & O'Connor, S.E. De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc. Natl. Acad. Sci. USA* **112**, 3205–3210 (2015).
72. Ma, S.M. *et al.* Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* **326**, 589–592 (2009).
73. Chiang, Y.M. *et al.* An efficient system for heterologous expression of secondary metabolite genes in *Aspergillus nidulans*. *J. Am. Chem. Soc.* **135**, 7720–7731 (2013).
74. Mutka, S.C., Carney, J.R., Liu, Y. & Kennedy, J. Heterologous production of epothilone C and D in *Escherichia coli*. *Biochemistry* **45**, 1321–1330 (2006).
75. Oßwald, C. *et al.* Modular construction of a functional artificial epothilone polyketide pathway. *ACS Synth. Biol.* **3**, 759–772 (2014).
76. Dueber, J.E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* **27**, 753–759 (2009).
77. Delebecque, C.J., Lindner, A.B., Silver, P.A. & Aldaye, F.A. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* **333**, 470–474 (2011).
78. Conrado, R.J. *et al.* DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic Acids Res.* **40**, 1879–1889 (2012).
79. Temme, K. *et al.* Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res.* **40**, 8773–8781 (2012).
80. Salis, H.M., Mirsky, E.A. & Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
81. Siegl, T., Tokovenko, B., Myronovskyi, M. & Luzhetskyy, A. Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. *Metab. Eng.* **19**, 98–106 (2013).

82. Ran, F.A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
83. Wang, H.H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
84. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
85. Bao, Z. *et al.* Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* **4**, 585–594 (2014).
86. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
87. Cobb, R.E., Wang, Y. & Zhao, H. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* **4**, 723–728 (2014).
88. Uguru, G.C. *et al.* Synthetic RNA silencing of actinorhodin biosynthesis in *Streptomyces coelicolor* A3(2). *PLoS ONE* **8**, e67509 (2013).
89. Na, D. *et al.* Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nat. Biotechnol.* **31**, 170–174 (2013).
90. Deb Roy, A., Grüschow, S., Cairns, N. & Goss, R.J. Gene expression enabling synthetic diversification of natural products: chemogenetic generation of pacidamycin analogs. *J. Am. Chem. Soc.* **132**, 12243–12245 (2010).
91. Arnison, P.G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 (2013).
92. Park, S.R., Park, J.W., Ban, Y.H., Sohng, J.K. & Yoon, Y.J. 2-Deoxystreptamine-containing aminoglycoside antibiotics: recent advances in the characterization and manipulation of their biosynthetic pathways. *Nat. Prod. Rep.* **30**, 11–20 (2013).
93. Kirby, J. & Keasling, J.D. Metabolic engineering of microorganisms for isoprenoid production. *Nat. Prod. Rep.* **25**, 656–661 (2008).
94. Xu, W., Gavia, D.J. & Tang, Y. Biosynthesis of fungal indole alkaloids. *Nat. Prod. Rep.* **31**, 1474–1487 (2014).
95. Schmidt, E.W. *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* **102**, 7315–7320 (2005).
96. Ruffner, D.E., Schmidt, E.W. & Heemstra, J.R. Assessing the combinatorial potential of the RiPP cyanobactin *tru* pathway. *ACS Synth. Biol.* **4**, 482–492 (2015).
97. Medema, M.H., van Raaphorst, R., Takano, E. & Breitling, R. Computational tools for the synthetic design of biochemical pathways. *Nat. Rev. Microbiol.* **10**, 191–202 (2012).
98. Birmingham, W.R. *et al.* Bioretrosynthetic construction of a didanosine biosynthetic pathway. *Nat. Chem. Biol.* **10**, 392–399 (2014).

Acknowledgments

We thank K. 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), the 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 US National Institutes of Health grants R01-CA127622 and R01-GM085770.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence should be addressed to B.S.M. or Y.J.Y.

ERRATUM

Reinvigorating natural product combinatorial biosynthesis with synthetic biology

Eunji Kim, Bradley S Moore & Yeo Joon Yoon

Nat. Chem. Biol. **11**, 649–659 (2015); published online 18 August 2015; corrected after print 23 September 2015

In the version of this article initially published, there were four typographical errors in the abstract and main text. The errors have been corrected in the HTML and PDF versions of the article.