Synthetic Biology of Natural Products

Rainer Breitling and Eriko Takano

Manchester Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM), Manchester Institute of Biotechnology, Faculty of Life Sciences, University of Manchester, Manchester M1 7DN, United Kingdom

Correspondence: rainer.breitling@manchester.ac.uk; eriko.takano@manchester.ac.uk

The diversity and natural modularity of their biosynthetic pathways has turned natural products into attractive, but challenging, targets for synthetic biology approaches. Here, we discuss the current state of the field, highlighting recent advances and remaining bottlenecks. Global genomic assessments of natural product biosynthetic capacities across large parts of microbial diversity provide a first survey of the available natural parts libraries and identify evolutionary design rules for further engineering. Methods for compound and pathway detection and characterization are developed increasingly on the basis of synthetic biology tools, contributing to an accelerated translation of genomic information into usable building blocks for pathway assembly. A wide range of methods is also becoming available for accessing ever larger parts of chemical space by rational diversification of natural products, guided by rapid progress in our understanding of the underlying biochemistry and enzymatic mechanisms. Enhanced genome assembly and editing tools, adapted to the needs of natural products research, facilitate the realization of ambitious engineering strategies, ranging from combinatorial library generation to high-throughput optimization of product titers. Together, these tools and concepts contribute to the emergence of a new generation of revitalized natural product research.

THE NATURAL PRODUCTS LANDSCAPE

The engineering of microbial systems for the enhanced production of natural products is one of the most obvious and popular applications of synthetic biology methodologies. Numerous excellent reviews, covering all aspects of the field, have been published during the last couple of years (Keasling 2008; Medema et al. 2011b; Cobb et al. 2013; Hranueli et al. 2013; Zakeri and Lu 2013; Cummings et al. 2014; Genilloud 2014; Luo et al. 2014; Porro et al. 2014; Quin and Schmidt-Dannert 2014; Singh

2014; Sleator 2014; Unkles et al. 2014; Wilson et al. 2014; Wright 2014; Breitling and Takano 2015). After a brief overview of the general ambitions and achievements of synthetic biology applications for natural products, we will then focus on selected aspects in which we either see major recent progress or important remaining bottlenecks.

Natural product production using engineered microorganisms is probably the application domain of synthetic biology in which its boundaries with classic genetic modification methods and traditional metabolic engineering

Editors: Daniel G. Gibson, Clyde A. Hutchison III, Hamilton O. Smith, and J. Craig Venter
Additional Perspectives on Synthetic Biology available at www.cshperspectives.org

Copyright © 2016 Cold Spring Harbor Laboratory Press; all rights reserved; doi: 10.1101/cshperspect.a023994
Cite this article as *Cold Spring Harb Perspect Biol* 2016;8:a023994

are most easily blurred. Nonetheless, following a number of highly successful case studies, it is considered one of the prime targets for biotechnological deployment of synthetic biology in the near future. Early commercialized examples range from the semisynthetic malaria drug artemisinin (University of California Berkeley/Amyris, Emeryville, CA/Sanofi, Paris) to the first consumer-market synthetic biology product, “natural” vanillin (Evolva, Reinach Switzerland/International Flavors and Fragrances, New York), both of them produced by fermentation in genetically modified yeast (Paddon and Keasling 2014; Kurita et al. 2015).

These success stories of synthetic biology have not only provided renewed motivation for the exploration of microbes as production systems for high-value compounds, they have also highlighted the challenges faced when moving from proof-of-principle heterologous synthesis of a biomolecule to its economically viable production on an industrial scale. The number and complexity of the necessary genetic modifications and the diversity of bottlenecks as a result of limited understanding of microbial physiology have been unexpected.

According to Porro et al. (2014), the necessary titer, yield, and productivity of microbial cell factories for industrial applications have rarely been achieved so far. Based on a thorough analysis of examples from primary metabolite production, they conclude that thermodynamic feasibility, and stoichiometric, redox, and energy balance, all need to be considered when assessing the economic feasibility of a microbial production system, both in terms of process efficiency and expected yield and purity. Moreover, they identify flux coupling and the introduction of irreversible reactions as important engineering targets to drive product generation. This is particularly important for natural products, for which the natural systems have typically not evolved toward high yields and titers, as minute amounts of bioactive molecules are all that is necessary for their biological function.

Optimizing all of these aspects has long been the aim of metabolic engineering, and synthetic biology offers additional opportunities for

achieving some of these goals. For example, Medema et al. (2011b) suggested that spatial engineering of biosynthetic pathways at various scales (from protein scaffolds to bacterial microcompartments to microbial consortia) can potentially be used to optimize product titers and increase metabolic pull by irreversibly trapping the desired end product in a compartment or vesicle, at the same time avoiding the potential toxicity seen for many natural products at high production levels. In particular, various protein-based bacterial microcompartments seem promising. For example, the Pdu microcompartment of *Citrobacter freundii* has already been engineered to encapsulate the pyruvate decarboxylase and alcohol dehydrogenase of *Zymomonas mobilis* to successfully produce ethanol in *Escherichia coli* (Lawrence et al. 2014; Chessher et al. 2015).

Another area of intense research is the discovery, modification, and overproduction of bioactive natural products from a variety of natural sources. In particular, the natural modularity of the biosynthetic pathways for some major classes of these bioactive molecules has attracted the interest of synthetic biologists, who see the potential of a “Lego-ization” of natural product biosynthesis. Two groups of natural products stand out in this respect, nonribosomal peptides and polyketides, both immensely diverse classes of compounds with a wide range of desirable bioactivities (from antibiotics and anticancer activities to cholesterol-lowering effects). Compounds in both of these groups are typically produced by large molecular assembly lines, which show natural modularity at multiple levels: enzymatic complexes responsible for the sequential incorporation of chemical units into a growing core compound are made up of individual catalytic domains in predictable patterns, and are, in turn, linked with different combinations of tailoring enzymes to yield the final product. Swapping of biosynthetic modules and domains and tailoring enzymes for the creation of chemical novelty has long been considered a promising strategy (often referred to as combinatorial biosynthesis), and recent advances in genome-editing capabilities have opened up the prospect of a plug-and-play ap-

proach to natural product discovery and production (Medema et al. 2011b; Frascch et al. 2013).

In the case of polyketide synthesis, the easy genome-based discovery (Helfrich et al. 2014) and extreme modularity of the system, and the resulting prospect of complete programmability, has led to predictions that around 20 of the 150 major commodity chemicals currently produced by the petrochemical industry could be synthesized by “mixing and matching” of naturally occurring polyketide synthase domains (Poust et al. 2014). Currently, only a small part of the theoretical chemical space accessible via polyketide synthesis is actually available through natural or engineered synthesis systems, but retrosynthetic approaches to predict and manipulate the specificity of polyketide synthase domains are constantly improving, and learning on a large scale from the successes and failures of repeated iterations of the design–build–test cycle of synthetic biology is expected to yield generalizable rules and principles for accessing ever larger fractions of chemical space. Some of the formidable challenges still faced in this domain, illustrating both the importance of synthetic biology tools and the present limitations of rational engineering of biochemical complexity, have recently been discussed by Cummings et al. (2014).

One additional exciting aspect of the synthetic biology approach is its systematic lowering of phylogenetic barriers. Traditionally, many complex natural products have been discovered in nonmodel organisms, such as plants and actinomycetes. Even today, industrial production processes often rely on native host strains that have been optimized through tedious rounds of mutagenesis and selection. Heterologous production was impeded by practical difficulties in genetically manipulating the huge biosynthetic gene clusters involved, especially when the extreme GC-rich nature of actinomycete genomes interfered with expression in model organisms, such as *E. coli*. Furthermore, extensive engineering of primary metabolism, providing essential precursors in sufficient amounts, were often required to allow production in a heterologous system. Synthetic biology promises to

change this. Not only is it becoming easier to refactor entire gene clusters for optimized activity in a nonnative environment, computational metabolic modeling and rapid genome editing make it possible to design production chassis with tailored primary metabolism quickly enough to be economically feasible (e.g., Song and Lee 2013). On the other hand, synthetic biology approaches are now used to streamline the genomes of actinomycete strains to obtain suitable generic hosts for a diverse range of secondary metabolites (Rebets et al. 2014), and computational modeling results indicate that a much broader range of species might have a metabolic network preadapted for the production of natural products (Zakrzewski et al. 2012). Moreover, the toolbox of synthetic biology, ranging from promoter libraries and expression vectors to reporter genes and regulatory circuitry, is rapidly growing, both in actinomycetes (Rebets et al. 2014; Weber et al. 2015b) and in a diverse range of other potential chassis organisms, including model species with a long track record of industrial application, such as yeast (Siddiqui et al. 2012) and *E. coli* (Li and Neubauer 2014).

Seyedsayamdost and Clardy (2014) have recently suggested the conceptual division of synthetic biology of natural products into three major areas. First, work on the “known knowns,” such as artemisinin or vanilla, in which the chemistry of the target compound, as well as its biosynthetic pathway, are well known and pathway engineering mainly aims at improved production, often in a more amenable host than the natural one. Second, research into the “known unknowns,” compounds whose existence is easily inferred from the analysis of bacterial genomes, but which have not yet been detected; in these cases, the contribution of synthetic biology is mostly foreseen in the awakening of the biosynthesis of cryptic metabolites, facilitating their chemical and functional characterization and, ultimately, their overproduction using the methods established for “known knowns.” And, finally, screen for the “unknown unknowns,” the large diversity of interesting molecules that are suspected to be still entirely undiscovered, belonging to new chemical classes not covered by current genome-based discovery

methods. As the continued discovery of new compounds and the associated biosynthetic enzyme activities forms the basis of the growing synthetic biology toolbox, we will start our discussion by exploring recent developments in the accelerated discovery of new natural products.

DIVERSITY OF NATURAL PRODUCTS

When the first genomic sequences of versatile secondary metabolite producers became available, it was soon discovered that a huge treasure trove of potential natural products had been missed by classic screening-based discovery strategies. Strains that were known to produce perhaps a handful of bioactive compounds were found to encode the machinery required for the biosynthesis of dozens of complex chemicals. Recently, a number of comprehensive analyses of the wealth of microbial genome data currently available have revealed the general patterns underlying this unexpected abundance of chemical diversity.

Cimermancic and colleagues (2014) annotated the genomes of 1154 microbial genomes across the entire prokaryotic section of the tree of life, using a computational algorithm that tries to minimize a bias toward the detection of known classes of biosynthetic gene clusters (BGCs). Using antiSMASH, 33,351 detected putative BGCs were annotated (Medema et al. 2011a; Blin et al. 2013; Weber et al. 2015a), and 10,724 of them were considered to be “high-confidence” clusters, as they either belonged to well-characterized classes or passed a manual inspection. Surprisingly, 40% of the putative BGCs turned out to encode for the biosynthesis of complex saccharides. This result was unexpected because this group of compounds was previously often neglected compared to known hyperdiverse classes, such as polyketides and nonribosomal peptides, despite the fact that bioactive saccharides have been reported before (Kersten et al. 2013). The function of most of the putative saccharides in the global analysis remains unknown, but they are clearly very widespread (93% of species contain at least one saccharide BGC and, in one-third of the species, they constitute the majority of the pu-

tative clusters) and surprisingly diverse (they are rarely shared even between closely related species, even less so than the notoriously diverse polyketides).

Even more strikingly, a global map of all high-confidence BGCs revealed a network containing large cliques of widely distributed BGC classes, which did not contain any experimentally characterized clusters. These orphan cliques, sometimes containing hundreds of members spread across a wide range of bacterial diversity, represent the most promising candidates for chemical and functional novelty in the expanded universe of natural product diversity. The most diverse of them, with a total of 1021 members across all subphyla of proteobacteria, as well as a range of other Gram-negative species, was shown experimentally by Cimermancic et al. (2014) to be responsible for the biosynthesis of aryl polyenes. The deep sequence divergence between cluster members and the broad, but discontinuous, range of the phylogenetic distribution indicates that the chemical diversity of this group of compounds, which has putative antioxidant activity, has been considerably underestimated before.

A study of the predicted biosynthetic gene clusters in 830 genome sequences of actinobacteria confirms the general pattern of astonishing diversity, detecting 11,422 putative BGCs in 4122 gene cluster families, each potentially encoding a different natural product scaffold. Extrapolating from this dataset to the complete diversity of actinobacteria, the investigators predict a total biosynthetic capacity of hundreds of thousands of natural products from this phylum alone, despite the fact that their analysis focused only on the “known unknowns,” that is, members of a handful of major well-characterized classes of secondary metabolites. The results of Cimermancic et al. (2014) indicate that this may still be an underestimate of the actual chemical potential of microbial metabolism.

The global assessment of biosynthetic capacity not only provides a map toward the identification of new building blocks for synthetic biology, it also may help guide the actual process of cluster engineering. A follow-up analysis of the same data set explored the patterns of gene

cluster evolution to identify rules that could help predict the interoperability of biosynthetic modules (Medema et al. 2014a). Observing how natural chemical diversity is created by evolution, they identified three specific constraints that should be taken into account when attempting cluster engineering for chemical novelty. First, their analysis delineates multienzyme subclusters that evolve as independent units in a process they describe as “bricks and mortar” evolution. These subclusters, rather than individual enzymes, could thus be considered as natural “biobrick” equivalents and would be the most promising starting point for a systematic recombination of biosynthetic functions for the assembly of new pathways. Second, they detect patterns of concerted evolution for groups of polyketide and nonribosomal peptide biosynthetic assembly lines, which have resulted in sets of sequence-homogenized catalytic domains; these groups are predicted to be particularly amenable to domain-shuffling strategies for diversity generation. And third, they highlight the fact that different classes of clusters show quite different evolutionary patterns with different levels of subcluster transfers, domain recruitment, concerted evolution, and internal duplication and diversification. Engineering strategies should take the specific evolutionary constraints revealed by this into account for optimized efficiency.

A more targeted genome survey by Yu et al. (2013) explored the diversity of phosphonate biosynthetic pathways using a signature gene (phosphoenolpyruvate mutase) as an indicator. This analysis revealed that phosphonates, which were previously considered biochemical oddities, are in fact widespread in microbes, the biosynthetic capacity being found in at least 5% of sequenced genomes from a diverse range of species. Further study of the genomic neighborhood of the signature genes suggests that the products of the phosphonate pathways are chemically very diverse (on the order of at least several hundred different chemical entities, as indicated by rarefaction analysis) and probably have a number of new and clinically relevant bioactivities. This analysis was later expanded to cover the entire predicted metabolome of ac-

tinomycetes, extrapolating to a potential chemical diversity of hundreds of thousands of possible drug leads encoded by members of this phylum alone (Doroghazi et al. 2014). Related approaches using signature genes in combination with high-throughput real-time polymerase chain reaction (PCR) have recently been suggested as an efficient means of strain prioritization for natural product discovery programs (Hindra et al. 2014), and their potential to identify “talented strains” was illustrated for diterpenoid-producing actinomycetes (Xie et al. 2014), albeit without a negative control, so that it remains possible that the successful discovery of numerous diterpenoids in one of the prioritized strains, *Streptomyces griseus* CB00830, is the result of the general giftedness of actinomycetes rather than a validation of the prioritization strategy.

One of the insights from genome-based natural product discovery was the realization that specialized metabolites can be found in a variety of unexpected sources. One of the underexploited sources of bioactive metabolites was recently highlighted by Sharon et al. (2014) who discuss the diverse set of metabolites produced by the human commensal microbiome and their potential effects on human physiology. As these metabolites have evolved in direct interaction with the human host, they are expected to be enriched in interesting drug leads, and metabolites from the microbiome turn out to play an important role in the development of cardiovascular disease, cancer, and other diseases, as well as influencing human behavior. Large-scale analysis of biosynthetic gene cluster complement of the human microbiome has already resulted in the discovery of a new thiopeptide antibiotic, lactocillin, from a member of the vaginal microbiome (Donia et al. 2014).

Another source of metabolic diversity is only recently becoming accessible. Advances in single-cell genome analysis allow, for the first time, the comprehensive exploration of the biosynthetic capacity of talented uncultivated bacteria, a rich source of potentially novel chemistry, as was shown, for example, by the diversity and uniqueness of biosynthetic gene clusters identified in the symbiotic bacteria of the ma-

rine sponge *Theonella swinhoei* (Wilson et al. 2014).

DISCOVERY OF NATURAL PRODUCTS

Although the genomic analyses help in the identification of building blocks and generic engineering strategies for the synthetic biology of natural products, chemistry-directed discovery is needed for turning “known unknowns” into “known knowns” as targets for drug development and biotechnology (Bologa et al. 2013). Luo et al. (2014) subdivide the field of natural products discovery into two main groups of approaches: top-down and bottom-up. The first, top-down, they consider to be the traditional primary means of discovery, using various perturbations to induce production of new compounds in their natural hosts without requiring prior knowledge of their biosynthetic machinery. The second class of approaches, in contrast, depends on genomic information and subsequent genetic manipulation to achieve the synthesis of a compound of interest, often in a heterologous host. Both of these strategies can be supported and accelerated by the techniques of synthetic biology, and progress in genomic and metabolomic technologies, as well as in bioinformatics, typically leads to rapid transition from top-down to bottom-up work once a compound of interest has first been detected.

Guo et al. (2015) introduced a strategy that combines a promoter reporter system that allows the color-based identification of strains overexpressing a previously cryptic target gene cluster with genome-scale random mutagenesis to create a library of genetic variants. They successfully applied this approach to the *pga* gene cluster of *Streptomyces* species PGA64, which had been recalcitrant to previous activation attempts, and identified two new anthroquinone aminoglycosides as the products associated with this gene cluster. In a complementary approach, Seyedsayamdost (2014) developed a high-throughput platform for awakening silent gene clusters using a library of small-molecule elicitors, avoiding the random mutagenesis required by the reporter-guided mutant selection strategy of Guo et al. (2015). In Seyedsayam-

dost's strategy, the activity of a silent gene cluster, for example, one that has been identified by genome analysis, is monitored by a targeted genetic reporter fusion, based on GFP or *lacZ* translationally fused to an essential gene of the cluster. Screening the resulting reporter strain against a library of bioactive compounds can rapidly identify chemical inducers of the silent cluster, as illustrated for the malleilactone and the burkholderac gene cluster. Once an efficient elicitor has been found, the induced product of the cluster can be detected and characterized, for example, by mass spectrometry-based methods. Multiplex strategies in which several clusters are monitored in parallel based on multiple reporters can easily be envisaged.

Computational methods for the detection of biosynthetic gene clusters in microbial genomes are continuously evolving, and the popular antiSMASH pipeline, which handles the majority of annotation tasks in one unified interface, is updated regularly (Medema et al. 2011a; Blin et al. 2013, 2014; Weber et al. 2015a). Like other genome annotation tools, antiSMASH relies on the availability of trustworthy reference annotations. The recently published community standard for the annotation of secondary metabolite biosynthetic gene clusters (MiBIG) provides a long-awaited tool for the rigorous documentation of the necessary information and has already been successfully applied in a large-scale standardized recuration of a vast body of legacy biochemical data on secondary metabolite biosynthesis (Medema et al. 2015). It is expected that, from now on, all major publications on the experimental characterization of biosynthetic gene clusters will be accompanied by MiBIG-compliant annotation data.

Many of the available tools, as well as other computational methods for the synthetic biology of natural products, were recently reviewed by Weber (2014) and Medema et al. (2012). The most interesting advances are provided by those tools that allow the direct integration of computational and experimental approaches. A recent example is Pep2Path (Medema et al. 2014b), a method for the mass spectrometry-guided mining for peptidic natural products. The Pep2Path software automates the process

of matching observed novel peptide metabolites to the most likely biosynthetic gene cluster using a rapid Bayesian probabilistic approach with high accuracy and precision. In combination with nanoDESI mass spectrometry (Watrous et al. 2013), this method can also be used for peptide discovery from environmental samples, matching observed compounds to related clusters in already sequenced genomes. Pep2Path is applicable to nonribosomal peptides (NRPs), as well as for ribosomally synthesized and post-translationally modified peptides (RiPPs), the latter being a group of compounds with unexpected prevalence in global genome analyses (Cimermancic et al. 2014). Methods for the rapid structural elucidation of NRPs and RiPPs by mass spectrometry have recently been developed (Mohimani et al. 2014a,b; Zhang et al. 2014), and a large-scale “metabologenomics” strategy matching bacterial genome sequences to mass spectrometry-based profiles of secreted metabolomes led to the discovery of a novel type of chlorinated nonribosomal peptide, tam-bromycin (Goering et al. 2016).

Although these approaches combine bottom-up and top-down discovery, progress is also made in pure bottom-up strategies. One example of this should suffice to illustrate the contribution of synthetic biology in this respect: Unkles et al. (2014) tackled the challenges of heterologous expression of silent fungal gene clusters by establishing a generic synthetic biology toolbox for this purpose. This includes improved in vivo cloning approaches, based on recombination in *Saccharomyces cerevisiae*; a talented host strain, *Aspergillus nidulans*, which is a native producer of natural products and therefore more likely to contain the necessary auxiliary machinery for successful biosynthesis; and the assembly of polycistronic genes using viral 2A peptide sequences to direct cotranslational cleavage, thus avoiding the requirements for individual marker selection systems and well-characterized promoter sequences for every single gene in a complex biosynthetic pathway. The application of this toolbox for the heterologous expression of the entire penicillin biosynthetic gene cluster of *Penicillium chrysogenum* from a single polycistronic gene showed

the convenience of this strategy. A recently developed platform for the rapid exchange of all native promoters in a target gene cluster with constitutively active promoters, based on yeast homologous recombination (Montiel et al. 2015), complements this system and further increases the speed with which silent or cryptic biosynthetic pathways can be awakened.

In related work, Ross et al. (2014) used synthetic biology tools to establish a platform for the capture and expression of biosynthetic gene clusters in *E. coli*. Using transformation-associated recombination in yeast they were able to capture targeted pathways without the usual size limitations associated with the traditional screening of large, randomized genomic libraries. Following the initial capture and heterologous expression of their proof-of-concept cluster, encoding the alterochromide pathway of *Pseudoalteromonas piscicida*, they proceeded to increase production, using λ -red recombination to bring the cluster under the control of a T7 promoter, together with the accessory phosphopantetheinyl transferase required for functionalizing the NRP synthase thiolation (T) domains. The product titers achieved were sufficient for detailed characterization of the biosynthetic pathway without further optimization.

All these approaches for the discovery of novel natural products will benefit from advances in functional screening methods that link chemical compounds to new bioactivities and modes of action, as illustrated by the proof-of-concept work of Kurita et al. (2015) on the functional annotation of natural product extracts by a combination of untargeted metabolomics and image-based phenotypic screening.

DESIGN OF NOVEL NATURAL PRODUCTS

Genome mining for new biosynthetic pathways and the discovery of new natural products from a variety of sources are only the first steps in a much more ambitious synthetic biology approach aiming at the generation of chemical novelties in the form of “nonnatural” natural products.

Beer et al. (2014) provided an important proof-of-concept advance in the engineering

of nonribosomal peptide biosynthesis. Targeting a single-module nonribosomal peptide synthase, IndC, which is responsible for the biosynthesis of the blue pigment indigoidin, they show successful exchange of the T domain of the enzyme, replacing it with natural and synthetic domains from different sources. In the process, they identified general “design principles” when considering the importance of the linker region between the T domain and its preceding adenylation domain. These principles and additional rules were incorporated in computer-aided design software that aims at helping users to create nonribosomal peptide synthases with any desired composition. Although the present experiments only aimed at exchanging T domains with the same substrate specificity, it is still an encouraging step forward (Calcott and Ackerley 2014 review some of the remaining challenges of a truly general NRP reengineering strategy). As a side note, it is amazing to realize that the major part of the work by Beer et al. (2014) was performed as part of a 6-month iGEM competition summer project—an achievement that illustrates that the synthetic biology of natural products has come a long way toward its goal of easy and predictable engineering of biological systems. Recent advances in understanding the molecular basis of the modularity of nonribosomal peptide synthesis, both at the level of the peptide synthases (Reimer et al. 2016) and at the level of precursor biosynthesis (Diez et al. 2015), will further expand our ability to create chemical novelty within this class in a predictable manner.

Sardar et al. (2015) advanced our ability to program the biosynthesis of RiPPs by refining our knowledge regarding the recognition sequences that direct processing enzymes to specific peptides. They show that, to a large extent, these recognition sequences are portable, and posttranslational modifications can be exchanged between RiPPs by importing the relevant enzyme and adding the corresponding recognition sequence to the precursor peptide, leading to diverse nonnatural products. As suggested by Medema et al. (2014a), Sardar et al. (2015) emphasize the importance of large libraries of RiPP pathways as a prerequisite for evolutionary analysis of sufficient breadth to

allow predictable engineering of the full scope of peptidic natural products.

In the area of polyketide biosynthesis, Sugimoto et al. (2014) managed for the first time to achieve full predictable reprogramming of a modular polyketide synthase to produce a product that was originally isolated in a different species. Using recombination and domain exchanges, they convert the aureothin biosynthetic pathway into a luteoreticuliculin assembly line, mimicking the natural evolutionary process of gene transfer, recombinations, and mutations. In the process, they generated “evolutionary” intermediate strains producing various luteoreticuliculin analogs, but also identified a number of unprecedented domain functions and context dependencies that reemphasize our need for deepened biochemical understanding of natural product biosynthetic machineries as a prerequisite for successful pathway engineering.

In another contribution to polyketide diversification, Riva et al. (2014) generated libraries of novel polyketide intermediates by feeding functionalized chemical probes to engineered bacterial strains that attach these probes to stalled biosynthetic intermediates that accumulate because of targeted mutations in the acyl carrier protein (ACP) domains of the native polyketide synthases. In their proof-of-concept study, they used this approach to generate a complex library of unnatural polyethers of potential therapeutic interest and suggest that similar libraries could be created by reinitiation of stalled biosynthesis on any ketosynthase, using diverse mimics of the natural malonyl-ACP extension intermediates. They also emphasize that the electron cryomicroscopy-based three-dimensional structure of a full-length polyketide synthase module (Dutta et al. 2014) will be a helpful guide for designing unnatural malonyl-ACP analogs with the best fit to all relevant active sites of the multienzyme.

Kung et al. (2014) take the structure-based approach one step further and use crystal structure information to guide the engineering of geranylgeranyl reductase for the controlled biosynthesis of diverse isoprenoids. They manage to achieve considerable control over the extent of reduction of the substrate, geranylgeranyl

pyrophosphate, GGPP (ranging from dihydro to hexahydro products). As GGPP is a key intermediate in the biosynthesis of an enormous range of isoprenoids (ranging from diterpenes and retinoids to carotenoids), this ability of controlled reduction opens up large new areas of chemical space for synthetic biology.

Another development that might have important implications for the rapid diversification of natural products is the increasing availability of cell-free platforms for synthetic biology (Toogood et al. 2015; Garamella et al. 2016). These systems allow the rapid prototyping, optimization, and combinatorial exploration of biosynthetic pathways, before introduction to a suitable chassis for overproduction.

BUILDING NATURAL PRODUCT PRODUCTION STRAINS

The engineering of natural product production strains obviously benefits from general advances in synthetic biology methodology for genome editing and assembly (Ellis et al. 2011; Esvelt and Wang 2013; Gibson 2014; Si et al. 2014; Selle and Barrangou 2015). However, recent years have also seen a number of advances specific to natural product research. For example, Cobb et al. (2014b) established an engineered CRISPR/Cas9 system for multiplex genome editing in *Streptomyces*, a genus notable for its abundant repertoire of secondary metabolites. The constructed pCRISPomyces plasmids are compatible with Golden Gate or isothermal assembly and are expected to make a major contribution to the large-scale genome editing for the purposes of awakening silent gene clusters, engineering and recombining pathways, as well as optimizing hosts strains. Other examples using the CRISPR system in actinomycetes have also been reported recently, making this groundbreaking technology widely applicable for this important group of natural product production hosts (Huang et al. 2015; Tong et al. 2015).

The necessary high-throughput DNA assembly methods able to deal with the challenges posed by the large and highly repetitive gene clusters for natural products have also advanced continuously, including methods facilitating the

domain swapping underlying combinatorial biosynthesis approaches toward creating natural product libraries (Cobb et al. 2014a). Methods based on site-specific recombination have recently been shown to provide further opportunities for the rapid assembly and reshuffling of metabolic pathways (Colloms et al. 2014). These methods have so far been shown on pathways of relatively limited size and complexity, but their potential to construct libraries of combinatorial pathway assemblies, with randomized gene order or ribosome binding site strength, as well as to modify pathways by facile addition or replacement of selected genes will be of considerable interest for natural product research.

The concept of combinatorial assembly has recently been taken to new extremes in the case of the *Klebsiella oxytoca* nitrogen fixation gene cluster, containing a total of 16 genes and 103 individual parts (open reading frames, promoters, ribosome binding sites, terminators) (Smanski et al. 2014). Starting from a comprehensively refactored and modularized version of the pathways and using a hierarchical DNA assembly strategy to construct libraries of variants with rearranged gene orders, orientation, and operon combinations, as well as simultaneous substitutions of the ribosome binding sites of all 16 genes, they achieved substantial improvements of cluster performance through multiple rounds of the design–build–test cycle of synthetic biology. Although the proof-of-concept target was a complex pathway of primary metabolism, the applicability of the same tools for natural product pathway engineering is obvious. The availability of integrating plasmids in actinobacteria, which are particularly talented producers of natural products, enables powerful multiplexed approaches for the combinatorial assembly of biosynthetic pathway variants (Fayed et al. 2015), which, in the future, could be applied not only for combinatorial biosynthesis, but also for more general refactoring strategies in these species.

CONCLUSION

It is obvious from the selected examples that the synthetic biology approaches to natural prod-

ucts research are progressing at a rapid and accelerating pace. Particularly noteworthy is the deepening integration between classical approaches to natural products and the new concepts of synthetic biology. This new, integrated community is quickly moving toward the critical mass that is necessary for translating numerous proof-of-concept studies and technological advances into an economically viable and generalizable strategy for the biotechnological industry, realizing the potential of natural products as drugs and chemicals for a wide variety of applications. It is also obvious from the preceding discussion that we are not quite there yet and that many challenges still remain to be addressed. We need more detailed knowledge of the engineering constraints imposed by enzymatic mechanisms and evolutionary history; we need more generic hosts optimized for rapid scale-up of production levels of natural products of interest; we need improved computational tools for pathway and host design; and we need even faster, more robust, and versatile genome assembly methods for increased throughput and improved predictability of the engineering process. If progress continues at the present rate, the synthetic biology of natural products is moving toward a bright future.

REFERENCES

- Beer R, Herbst K, Ignatiadis N, Kats I, Adlung L, Meyer H, Niopek D, Christiansen T, Georgi F, Kurzawa N, et al. 2014. Creating functional engineered variants of the single-module non-ribosomal peptide synthetase IndC by T domain exchange. *Mol Biosyst* **10**: 1709–1718.
- Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T. 2013. antiSMASH 2.0—A versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res* **41**: W204–W212.
- Blin K, Kazempour D, Wohlleben W, Weber T. 2014. Improved lanthipeptide detection and prediction for antiSMASH. *PLoS ONE* **9**: e89420.
- Bologa CG, Ursu O, Oprea TI, Melancon CE III, Tegos GP. 2013. Emerging trends in the discovery of natural product antibacterials. *Curr Opin Pharmacol* **13**: 678–687.
- Breitling R, Takano E. 2015. Synthetic biology advances for pharmaceutical production. *Curr Opin Biotechnol* **35C**: 46–51.
- Calcott MJ, Ackerley DF. 2014. Genetic manipulation of non-ribosomal peptide synthetases to generate novel bioactive peptide products. *Biotechnol Lett* **36**: 2407–2416.
- Chessher A, Breitling R, Takano E. 2015. Bacterial micro-compartments: Biomaterials for synthetic biology-based compartmentalization strategies. *ACS Biomater Sci Eng* **1**: 345–351.
- Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, et al. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**: 412–421.
- Cobb RE, Luo Y, Freestone T, Zhao H. 2013. Drug discovery and development via synthetic biology. In *Synthetic biology*, pp. 183–206. Elsevier, Amsterdam.
- Cobb RE, Ning JC, Zhao H. 2014a. DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. *J Ind Microbiol Biotechnol* **41**: 469–477.
- Cobb RE, Wang Y, Zhao H. 2014b. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth Biol* **4**: 723–728.
- Colloms SD, Merrick CA, Olorunniji FJ, Stark WM, Smith MCM, Osbourn A, Keasling JD, Rosser SJ. 2014. Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res* **42**: e23.
- Cummings M, Breitling R, Takano E. 2014. Steps towards the synthetic biology of polyketide biosynthesis. *FEMS Microbiol Lett* **351**: 116–125.
- Diez V, Loznik M, Taylor S, Winn M, Rattray NJ, Podmore H, Micklefield J, Goodacre R, Medema MH, Muller U, et al. 2015. Functional exchangeability of oxidase and dehydrogenase reactions in the biosynthesis of hydroxyphenylglycine, a nonribosomal peptide building block. *ACS Synth Biol* **4**: 796–807.
- Donia MS, Cimermancic P, Schulze CJ, Wieland Brown LC, Martin J, Mitreva M, Clardy J, Linington RG, Fischbach MA. 2014. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**: 1402–1414.
- Doroghazi JR, Albright JC, Goering AW, Ju KS, Haines RR, Tchalukov KA, Labeda DP, Kelleher NL, Metcalf WW. 2014. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat Chem Biol* **10**: 963–968.
- Dutta S, Whitcher JR, Hansen DA, Hale WA, Chemler JA, Congdon GR, Narayan AR, Hakansson K, Sherman DH, Smith JL, et al. 2014. Structure of a modular polyketide synthase. *Nature* **510**: 512–517.
- Ellis T, Adie T, Baldwin GS. 2011. DNA assembly for synthetic biology: From parts to pathways and beyond. *Integr Biol (Cambr)* **3**: 109–118.
- Esvelt KM, Wang HH. 2013. Genome-scale engineering for systems and synthetic biology. *Mol Syst Biol* **9**: 641.
- Fayed B, Ashford DA, Hashem AM, Amin MA, El Gazayerly ON, Gregory MA, Smith MC. 2015. Multiplexed integrating plasmids for engineering of the erythromycin gene cluster for expression in *Streptomyces* spp. and combinatorial biosynthesis. *Appl Environ Microbiol* **81**: 8402–8413.
- Frasch HJ, Medema MH, Takano E, Breitling R. 2013. Design-based re-engineering of biosynthetic gene clusters:

- Plug-and-play in practice. *Curr Opin Biotechnol* **24**: 1144–1150.
- Garamella J, Marshall R, Rustad M, Noireaux V. 2016. The all *E. coli* TX-TL toolbox 2.0: A platform for cell-free synthetic biology. *ACS Synth Biol* **5**: 344–355.
- Genilloud O. 2014. The re-emerging role of microbial natural products in antibiotic discovery. *Antonie Van Leeuwenhoek* **106**: 173–188.
- Gibson DG. 2014. Programming biological operating systems: Genome design, assembly and activation. *Nat Methods* **11**: 521–526.
- Goering AW, McClure RA, Doroghazi JR, Albright JC, Haverland NA, Zhang Y, Ju K-S, Thomson RJ, Metcalf WW, Kelleher NL. 2016. Metabologenomics: Correlation of microbial gene clusters with metabolites drives discovery of a nonribosomal peptide with an unusual amino acid monomer. *ACS Cent Sci* **2**: 99–108.
- Guo F, Xiang S, Li L, Wang B, Rajasarkka J, Grondahl-Yli-Hannuksela K, Ai G, Metsä-Ketela M, Yang K. 2015. Targeted activation of silent natural product biosynthesis pathways by reporter-guided mutant selection. *Metabolic Eng* **28**: 134–142.
- Helfrich EJN, Reiter S, Piel J. 2014. Recent advances in genome-based polyketide discovery. *Curr Opin Biotechnol* **29**: 107–115.
- Hindra, Huang T, Yang D, Rudolf JD, Xie P, Xie G, Teng Q, Lohman JR, Zhu X, Huang Y, et al. 2014. Strain prioritization for natural product discovery by a high-throughput real-time PCR method. *J Nat Prod* **77**: 2296–2303.
- Hranueli D, Starcevic A, Zucko J, Rojas JD, Diminic J, Baranasic D, Gacesa R, Padilla G, Long PF, Cullum J. 2013. Synthetic biology: A novel approach for the construction of industrial microorganisms. *Food Technol Biotechnol* **51**: 3–11.
- Huang H, Zheng G, Jiang W, Hu H, Lu Y. 2015. One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta Biochim Biophys Sinica* **47**: 231–243.
- Keasling JD. 2008. Synthetic biology for synthetic chemistry. *ACS Chem Biol* **3**: 64–76.
- Kersten RD, Ziemert N, Gonzalez DJ, Duggan BM, Nizet V, Dorrestein PC, Moore BS. 2013. Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. *Proc Natl Acad Sci* **110**: E4407–E4416.
- Kung Y, McAndrew RP, Xie X, Liu CC, Pereira JH, Adams PD, Keasling JD. 2014. Constructing tailored isoprenoid products by structure-guided modification of geranylgeranyl reductase. *Structure* **22**: 1028–1036.
- Kurita KL, Glassey E, Linington RG. 2015. Integration of high-content screening and untargeted metabolomics for comprehensive functional annotation of natural product libraries. *Proc Natl Acad Sci* **112**: 11999–12004.
- Lawrence AD, Frank S, Newnham S, Lee MJ, Brown IR, Xue WF, Rowe ML, Mulvihill DP, Prentice MB, Howard MJ, et al. 2014. Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS Synth Biol* **3**: 454–465.
- Li J, Neubauer P. 2014. *Escherichia coli* as a cell factory for heterologous production of nonribosomal peptides and polyketides. *New Biotechnol* **31**: 579–585.
- Luo Y, Cobb RE, Zhao H. 2014. Recent advances in natural product discovery. *Curr Opin Biotechnol* **30**: 230–237.
- Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. 2011a. antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* **39**: W339–W346.
- Medema MH, Breitling R, Bovenberg R, Takano E. 2011b. Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. *Nat Rev Microbiol* **9**: 131–137.
- Medema MH, van Raaphorst R, Takano E, Breitling R. 2012. Computational tools for the synthetic design of biochemical pathways. *Nat Rev Microbiol* **10**: 191–202.
- Medema MH, Cimermancic P, Sali A, Takano E, Fischbach MA. 2014a. A systematic computational analysis of biosynthetic gene cluster evolution: Lessons for engineering biosynthesis. *PLoS Comput Biol* **10**: e1004016.
- Medema MH, Paalvast Y, Nguyen DD, Melnik A, Dorrestein PC, Takano E, Breitling R. 2014b. Pep2Path: Automated mass spectrometry-guided genome mining of peptidic natural products. *PLoS Comput Biol* **10**: e1003822.
- Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC, et al. 2015. Minimum information about a biosynthetic gene cluster. *Nat Chem Biol* **11**: 625–631.
- Mohimani H, Kersten RD, Liu WT, Wang M, Purvine SO, Wu S, Brewer HM, Pasa-Tolic L, Bandeira N, Moore BS, et al. 2014a. Automated genome mining of ribosomal peptide natural products. *ACS Chem Biol* **9**: 1545–1551.
- Mohimani H, Liu WT, Kersten RD, Moore BS, Dorrestein PC, Pevzner PA. 2014b. NRPquest: Coupling mass spectrometry and genome mining for nonribosomal peptide discovery. *J Nat Prod* **77**: 1902–1909.
- Montiel D, Kang HS, Chang FY, Charlop-Powers Z, Brady SE. 2015. Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. *Proc Natl Acad Sci* **112**: 8953–8958.
- Paddon CJ, Keasling JD. 2014. Semi-synthetic artemisinin: A model for the use of synthetic biology in pharmaceutical development. *Nat Rev Microbiol* **12**: 355–367.
- Porro D, Branduardi P, Sauer M, Mattanovich D. 2014. Old obstacles and new horizons for microbial chemical production. *Curr Opin Biotechnol* **30**: 101–106.
- Poust S, Hagen A, Katz L, Keasling JD. 2014. Narrowing the gap between the promise and reality of polyketide synthases as a synthetic biology platform. *Curr Opin Biotechnol* **30**: 32–39.
- Quin MB, Schmidt-Dannert C. 2014. Designer microbes for biosynthesis. *Curr Opin Biotechnol* **29**: 55–61.
- Rebets Y, Brötz E, Tokovenko B, Luzhetskyy A. 2014. Actinomycetes biosynthetic potential: How to bridge in silico and in vivo? *J Ind Microbiol Biotechnol* **41**: 387–402.
- Reimer JM, Aloise MN, Harrison PM, Schmeing TM. 2016. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature* **529**: 239–242.

- Riva E, Wilkening I, Gazzola S, Li WM, Smith L, Leadlay PF, Tosin M. 2014. Chemical probes for the functionalization of polyketide intermediates. *Angew Chem Int Ed Engl* **53**: 11944–11949.
- Ross AC, Gulland LE, Dorrestein PC, Moore BS. 2014. Targeted capture and heterologous expression of the *Pseudomonas* Alterochromide gene cluster in *Escherichia coli* represents a promising natural product exploratory platform. *ACS Synth Biol* **4**: 414–420.
- Sardar D, Pierce E, McIntosh JA, Schmidt EW. 2015. Recognition sequences and substrate evolution in cyanobactin biosynthesis. *ACS Synth Biol* **4**: 167–176.
- Selle K, Barrangou R. 2015. Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends Microbiol* **23**: 225–232.
- Seyedsayamdost MR. 2014. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc Natl Acad Sci* **111**: 7266–7271.
- Seyedsayamdost MR, Clardy J. 2014. Natural products and synthetic biology. *ACS Synth Biol* **3**: 745–747.
- Sharon G, Garg N, Debelius J, Knight R, Dorrestein PC, Mazmanian SK. 2014. Specialized metabolites from the microbiome in health and disease. *Cell Metab* **20**: 719–730.
- Si T, Xiao H, Zhao H. 2014. Rapid prototyping of microbial cell factories via genome-scale engineering. *Biotechnol Adv* **33**: 1420–1432.
- Siddiqui MS, Thodey K, Trenchard I, Smolke CD. 2012. Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Res* **12**: 144–170.
- Singh V. 2014. Recent advancements in synthetic biology: Current status and challenges. *Gene* **535**: 1–11.
- Sleator RD. 2014. The synthetic biology future. *Bioengineered* **5**: 69–72.
- Smanski MJ, Bhatia S, Zhao D, Park Y, Woodruff L BA, Giannoukos G, Ciulla D, Busby M, Calderon J, Nicol R, et al. 2014. Functional optimization of gene clusters by combinatorial design and assembly. *Nat Biotechnol* **32**: 1241–1249.
- Song CW, Lee SY. 2013. Rapid one-step inactivation of single or multiple genes in *Escherichia coli*. *Biotechnol J* **8**: 776–784.
- Sugimoto Y, Ding L, Ishida K, Hertweck C. 2014. Rational design of modular polyketide synthases: Morphing the aureothin pathway into a luteoreticulic assembly line. *Angew Chem Int Ed Engl* **53**: 1560–1564.
- Tong Y, Charusanti P, Zhang L, Weber T, Lee SY. 2015. CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth Biol* **4**: 1020–1029.
- Toogood HS, Ni Cheallaigh A, Tait S, Mansell DJ, Jervis A, Lygidakis A, Humphreys L, Takano E, Gardiner JM, Scrutton NS. 2015. Enzymatic menthol production: One-pot approach using engineered *Escherichia coli*. *ACS Synth Biol* **4**: 1112–1123.
- Unkles SE, Valiante V, Mattern DJ, Brakhage AA. 2014. Synthetic biology tools for bioprospecting of natural products in eukaryotes. *Chem Biol* **21**: 502–508.
- Watrous J, Roach P, Heath B, Alexandrov T, Laskin J, Dorrestein PC. 2013. Metabolic profiling directly from the Petri dish using nanospray desorption electrospray ionization imaging mass spectrometry. *Anal Chem* **85**: 10385–10391.
- Weber T. 2014. In silico tools for the analysis of antibiotic biosynthetic pathways. *Int J Med Microbiol* **304**: 230–235.
- Weber T, Blin K, Duddela S, Krug D, Kim HU, Brucoleri R, Lee SY, Fischbach MA, Muller R, Wohlleben W, et al. 2015a. antiSMASH 3.0—A comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* **43**: W237–W243.
- Weber T, Charusanti P, Musiol-Kroll EM, Jiang X, Tong Y, Kim HU, Lee SY. 2015b. Metabolic engineering of antibiotic factories: New tools for antibiotic production in actinomycetes. *Trends Biotechnol* **33**: 15–26.
- Wilson MC, Mori T, Ruckert C, Uria AR, Helf MJ, Takada K, Gernert C, Steffens UA, Heycke N, Schmitt S, et al. 2014. An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**: 58–62.
- Wright G. 2014. Perspective: Synthetic biology revives antibiotics. *Nature* **509**: S13.
- Xie P, Ma M, Rateb ME, Shaaban KA, Yu Z, Huang SX, Zhao LX, Zhu X, Yan Y, Peterson RM, et al. 2014. Biosynthetic potential-based strain prioritization for natural product discovery: A showcase for diterpenoid-producing actinomycetes. *J Nat Prod* **77**: 377–387.
- Yu X, Doroghazi JR, Janga SC, Zhang JK, Circello B, Griffin BM, Labeda DP, Metcalf WW. 2013. Diversity and abundance of phosphonate biosynthetic genes in nature. *Proc Natl Acad Sci* **110**: 20759–20764.
- Zakeri B, Lu TK. 2013. Synthetic biology of antimicrobial discovery. *ACS Synth Biol* **2**: 358–372.
- Zakrzewski P, Medema MH, Gevorgyan A, Kierzek AM, Breitling R, Takano E. 2012. MultiMetEval: Comparative and multi-objective analysis of genome-scale metabolic models. *PLoS ONE* **7**: e51511.
- Zhang Q, Ortega M, Shi Y, Wang H, Melby JO, Tang W, Mitchell DA, van der Donk WA. 2014. Structural investigation of ribosomally synthesized natural products by hypothetical structure enumeration and evaluation using tandem MS. *Proc Natl Acad Sci* **111**: 12031–12036.