Natural Products and Synthetic Biology: Where We Are and Where We Need To Go

Sylvia Kunakom,[a] Alessandra S. Eustáquio[a]

[a]Department of Medicinal Chemistry and Pharmacognosy and Center for Biomolecular Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, Illinois, USA

ABSTRACT The biosynthetic talent of microorganisms has been harnessed for drug discovery for almost a century. Microbial metabolites not only account for the majority of antibiotics available today, but have also led to anticancer, immunosuppressant, and cholesterol-lowering drugs. Yet, inherent challenges of natural products—including inadequate supply and difficulties with structure diversification—contributed to their deprioritization as a source of pharmaceuticals. In recent years, advances in genome sequencing and synthetic biology spurred a renewed interest in natural products. Bacterial genomes encode an abundance of natural products awaiting discovery. Synthetic biology can facilitate not only discovery and improvements in supply, but also structure diversification. This perspective highlights prior accomplishments in the field of synthetic biology and natural products by the scientific community at large, including research from our laboratory. We also provide our opinion as to where we need to go to continue advancing the field.

KEYWORDS drug discovery, bacterial genomes, biosynthesis, metabolites, structure diversification

Drug discovery starts with screening libraries of compounds in biological assays of interest. The input in these screens can be synthetic compounds or natural products: that is, extracts obtained from living systems or pure compounds. Natural products provide privileged scaffolds for drug discovery as they offer structural complexity and physicochemical characteristics that evolved to interact with biological systems. Yet, the natural products themselves rarely become drugs. Rather, they often serve as starting points for drug discovery where natural product derivatives with improved pharmacokinetics and pharmacodynamics properties are then developed into drugs (1, 2).

The structural complexity of natural products is a double-edged sword. Natural products occupy areas of chemical space that are distinct from synthetic molecules, increasing the likelihood of target interaction and hit identification when natural products are included in drug discovery screens (1). At the same time, their structural complexity reduces synthetic tractability, making structure diversification challenging. Moreover, the frequent rediscovery of known compounds and inadequate supply also contributed to deprioritization of natural products research by the private sector starting in the 1990s (3).

In bacteria, biosynthesis of natural products is often encoded in genes that are colocalized in the genome, forming what are referred to as biosynthetic gene clusters (BGCs). About 100,000 bacterial genomes have now been sequenced, and this number will continue to increase as massive genomic and metagenomic efforts are under way. Bioinformatics tools have been developed for automated BGC identification (e.g., see reference 4), leading to an abundance of orphan BGCs for which the encoded com-
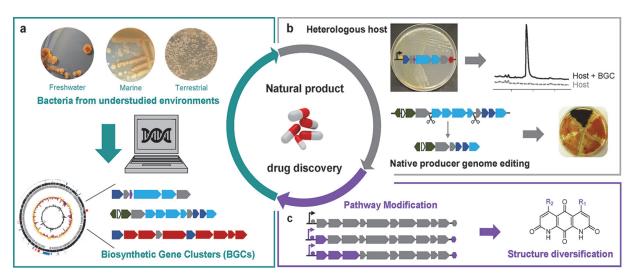
FIG 1 Genomic, genetic, and synthetic biology approaches toward natural product drug discovery. (a) Exploration of bacteria from understudied environments and taxa follows the hypothesis that different selective pressures may lead to the evolution of distinct chemistry. Genomes are sequenced and biosynthetic gene clusters (BGCs) are identified using bioinformatic tools. (b) BGCs of interest are selected. Heterologous expression in suitable hosts can be used to streamline natural product discovery. Genome editing of native producers offers an alternative approach to aid discovery and to study gene and BGC function. (c) Pathway modification via BGC reprogramming contributes to structure diversification of natural products.

pounds are unknown. Given that only a small fraction of known natural products have been connected to BGCs, BGC dereplication (that is, assignment of BGCs to known compounds) helps with genome mining approaches by identifying BGCs that are likely to encode novel natural products (5). BGC dereplication still leaves a wealth of unknown BGCs, and the challenge now is how to prioritize BGCs for discovery (6). Regardless of which criteria we use to select BGCs, genome mining will undoubtedly contribute to natural product discovery in the years to come. We use a broad definition of genome mining here to mean mining genomes not only to identify BGCs and obtain the encoded compounds, but also to connect known compounds to their BGCs.

One approach we (7) and others (e.g., reference 8) are following to counteract rediscovery and to try to expand the diversity of natural product scaffolds available for drug discovery is to explore understudied environments, the hypothesis being that different environments offer distinct selective pressures that may lead to the evolution of structurally diverse natural products (Fig. 1a). Synthetic biology can facilitate such natural product discovery efforts as it enables going from the DNA sequence in a computer to pure natural products in a bottle without ever having to work with the organisms that encode the information. This can be done by synthesizing BGCs followed by heterologous expression in a host organism (Fig. 1b). Circumventing native producers is relevant because it allows us, for example, to obtain compounds from organisms that have yet to be cultured or to obtain compounds from metagenomic data sets (9). In addition, for cultured organisms, it provides a more streamlined way to increase yields and engineer biosynthesis because native producers may not be genetically tractable.

We would be remiss not to mention that we actually believe that any organism can be eventually genetically engineered and that it is a matter of how much time and effort one wants to put into attempting to develop genetics for organisms that contain BGCs of interest. In fact, we are interested in developing reverse genetics methods (7, 10), as reverse genetics allows us to probe the function of genes in their native setting (Fig. 1b). Reverse genetics also allows us to answer questions such as the function of natural products for the producing organism, questions we are interested in from a basic science perspective. Yet, if the goal is to increase the number of known natural products available to the scientific community in a high-throughput manner, then heterologous expression of BGCs in well-characterized hosts offers the advantage of

speed. In this regard, the synthetic biology community seems to be moving away from the idea that "one host fits all" to have hosts tailored to the source of natural products and the compound class. In addition to genetic engineering of native producers, host development is an active area of investigation in our laboratory (Fig. 1b).

Synthetic biology can also complement synthesis efforts and contribute to structure diversification (Fig. 1c). Notable examples include reprogramming of polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) modular enzymes. Although the production of natural product derivatives via PKS and NRPS engineering has been demonstrated, engineered systems often suffer from low yields. The Abe group (11) has recently shown that an evolution-inspired engineering strategy can lead to the production of natural product derivatives in good yields. The authors took knowledge regarding the evolution of PKS and NRPS systems into consideration before engineering ring contraction, ring expansion, and alkyl chain diversification of a family of cyclic depsipeptides. Another class of natural products that offer great potential for engineered biosynthesis is ribosomally synthesized and posttranslationally modified peptides (RiPPs). RiPPs have the advantage of being encoded in relatively small BGCs, which greatly facilitates DNA synthesis and synthetic biology efforts toward structure diversification. Most importantly, the broad substrate tolerance of RiPP biosynthetic enzymes allows processing of variable core sequences. A nice demonstration of exploring the promiscuity of RiPP biosynthetic enzymes to discover new inhibitors of drug targets was recently provided by the van der Donk group (12). In this study, the authors coupled a plasmid-encoded library of lanthipeptides with a two-hybrid system to identify inhibitors of a protein-protein interaction critical to the HIV infection cycle. Synthetic biology efforts also benefit from increased knowledge of natural product biosynthesis (10) and the discovery and characterization of enzymes (13) that can be used in pathway construction for engineered biosynthesis.

We envision that in the near future the natural products community will have access to a comprehensive synthetic biology tool box, including computational tools for pathway design, wet lab tools for pathway construction, and an assortment of hosts for BGC expression and compound production. Computational tools are already getting increasingly sophisticated. For example, ClusterCAD, developed by the Keasling group (14), offers an exciting web-based platform to assist with PKS engineering. Moreover, reading DNA is still orders of magnitude cheaper than writing it. Therefore, the field will greatly benefit from cost reduction in DNA synthesis. Recent advances in enzymatic oligonucleotide synthesis (15) may eventually result in new methods that overcome the size limitation ($\sim$200 nucleotides) of current phosphoramidite-based synthesis, which can in turn reduce the reliance on assembly and facilitate the synthesis of BGCs (sizes can range from $\sim$5 to 200 kb). However, for enzymatic synthesis to become practical, many challenges remain to be overcome, such as improvements in yields and demonstration that large oligonucleotides can be generated with accuracy (15). How long it will take for DNA synthesis costs to significantly drop and catch up with sequencing remains to be seen. Yet, because DNA synthesis is one of the major limitations of synthetic biology, we expect a commensurate effort by the private and academic sectors will lead to innovative approaches and significant cost reduction in the next several years.

Furthermore, efficient *in vitro* and *in vivo* methods to assemble large, repetitive DNA sequences with ease are necessary to leverage the modular logic of PKSs and NRPSs and expand natural product diversity. As we explore untapped microbial taxa, heterologous hosts that work well for these taxa are also necessary. Additionally, because drug discovery often requires chemical transformations that are not necessarily found in nature, directed evolution (16) of enzymes and pathways is expected to greatly contribute to synthetic biology efforts aimed at structure diversification of natural products.

Synthetic biology is still a trial-and-error endeavor relying on build-test-learn cycles. We still need to learn how to design pathways that work efficiently. Or do we? We made the mistake before to presume that we could design enzymes, or, in other words, that

we could predict function based on sequence. Frances Arnold (16) and others showed us that directing evolution via random mutagenesis and selection/screening is actually the more rational way to go, because we are, as of yet, unable to predict beneficial mutations. Could the same be true for the design of chimeric, modular enzymes or for biosynthetic pathway design? In any case, we expect that evolution-inspired approaches such as those recently reported by the Abe group (11) have the best chances of leading to the desired function, and we may one day learn how to design by observing evolution.

ACKNOWLEDGMENTS

REFERENCES

1. Huffman B, Shenvi RA. 2019. Natural products in the "marketplace": interfacing synthesis and biology. J Am Chem Soc 141:3332–3346. https://doi.org/10.1021/jacs.8b11297.

2. Newman DJ, Cragg GM. 2016. Natural products as sources of new drugs from 1981 to 2014. J Nat Prod 79:629–661. https://doi.org/10.1021/acs.jnatprod.5b01055.

3. Carter GT. 2011. Natural products and Pharma 2011: strategic changes spur new opportunities. Nat Prod Rep 28:1783–1789. https://doi.org/10.1039/c1np00033k.

4. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res 45:W36–W41. https://doi.org/10.1093/nar/gkx319.

5. Dejong CA, Chen GM, Li H, Johnston CW, Edwards MR, Rees PN, Skinnider MA, Webster AL, Magarvey NA. 2016. Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. Nat Chem Biol 12:1007–1014. https://doi.org/10.1038/nchembio.2188.

6. Ziemert N, Alanjary M, Weber T. 2016. The evolution of genome mining in microbes—a review. Nat Prod Rep 33:988–1005. https://doi.org/10.1039/C6NP00025H.

7. Braesel J, Crnkovic CM, Kunstman KJ, Green SJ, Maienschein-Cline M, Orjala J, Murphy BT, Eustáquio AS. 2018. Complete genome of Micromonospora sp. strain B006 reveals biosynthetic potential of a Lake Michigan actinomycete. J Nat Prod 81:2057–2068. https://doi.org/10.1021/acs.jnatprod.8b00394.

8. Chevrette MG, Carlson CM, Ortega HE, Thomas C, Ananiev GE, Barns KJ, Book AJ, Cagnazzo J, Carlos C, Flanigan W, Grubbs KJ, Horn HA, Hoffmann FM, Klassen JL, Knack JJ, Lewin GR, McDonald BR, Muller L, Melo WGP, Pinto-Tomas AA, Schmitz A, Wendt-Pienkowski E, Wildman S, Zhao M, Zhang F, Bugni TS, Andes DR, Pupo MT, Currie CR. 2019. The antimi-crobial potential of Streptomyces from insect microbiomes. Nat Commun 10:516. https://doi.org/10.1038/s41467-019-08438-0.

9. Wu C, Shang Z, Lemetre C, Ternei MA, Brady SF. 2019. Cadasides, calcium-dependent acidic lipopeptides from the soil metagenome that are active against multidrug resistant bacteria. J Am Chem Soc 141:3910–3919. https://doi.org/10.1021/jacs.8b12087.

10. Braesel J, Lee J-H, Arnould B, Murphy BT, Eustáquio AS. 2019. Diazaqui-nomycin biosynthetic gene clusters from marine and freshwater actinomycetes. J Nat Prod 82:937–946. https://doi.org/10.1021/acs.jnatprod.8b01028.

11. Awakawa T, Fujioka T, Zhang L, Hoshino S, Hu Z, Hashimoto J, Kozone I, Ikeda H, Shin-Ya K, Liu W, Abe I. 2018. Reprogramming of the antimycin NRPS-PKS assembly lines inspired by gene evolution. Nat Commun 9:3534. https://doi.org/10.1038/s41467-018-05877-z.

12. Yang X, Lennard KR, He C, Walker MC, Ball AT, Doigneaux C, Tavassoli A, van der Donk WA. 2018. A lanthipeptide library used to identify a protein-protein interaction inhibitor. Nat Chem Biol 14:375–380. https://doi.org/10.1038/s41589-018-0008-5.

13. Davis TD, Kunakom S, Burkart MD, Eustáquio AS. 2018. Preparation, assay, and application of chlorinase SalL for the chemoenzymatic synthesis of S-adenosyl-L-methionine and analogs. Methods Enzymol 604:367–388. https://doi.org/10.1016/bs.mie.2018.02.012.

14. Eng CH, Backman TWH, Bailey CB, Magnan C, Garcia Martin H, Katz L, Baldi P, Keasling JD. 2018. ClusterCAD: a computational platform for type I modular polyketide synthase design. Nucleic Acids Res 46:D509–D515. https://doi.org/10.1093/nar/gkx893.

15. Palluk S, Arlow DH, de Rond T, Barthel S, Kang JS, Bector R, Baghdassarian HM, Truong AN, Kim PW, Singh AK, Hillson NJ, Keasling JD. 2018. De novo DNA synthesis using polymerase-nucleotide conjugates. Nat Biotechnol 36:645–650. https://doi.org/10.1038/nbt.4173.

16. Arnold FH. 2018. Directed evolution: bringing new chemistry to life. Angew Chem Int Ed Engl 57:4143–4148. https://doi.org/10.1002/anie.201708408.