ACS SyntheticBiology

Viewpoint

pubs.acs.org/synthbio

Natural Products and Synthetic Biology

Mohammad R. Seyedsayamdost and Jon Clardy*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, United States

> *There are known knowns; these are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns - the ones we don't know we don't know.*
>
> Donald Rumsfeld, Former U.S. Secretary of Defense, 2/12/2002

While the phrasing of this tripartite classification was not intended to describe the current state of natural products research, it could have been, as it provides an apt description of the field today. The organic molecules lumped under the term natural products elude an easy definition. They are a structurally diverse collection of molecules produced by organisms in an idiosyncratic fashion, and their biological activities reflect the diversity of their producers and structures. Inclusion in or exclusion from the natural product family reflects the elasticity of classification. Consider three molecules that are closely related structurally and biosynthetically: tryptophan, serotonin, and psilocin (Figure 1a). Tryptophan's universal distribution as a proteinogenic amino acid excludes it from being a natural product. Serotonin's widespread distribution and well-understood biological activities result in its classification as a hormone or neurotransmitter, while psilocin, a hallucinogenic relative produced by only a few New World mushrooms, is a natural product. All natural products have one feature in common, biosyntheses by a genetically encoded pathway, and the possibilities for re-engineering these pathways connects natural products to synthetic biology.

The opening quotation's classification would differentiate natural products into the known knowns, such as the antibiotic erythromycin, for which we know the molecule (and in this case its genetically encoded biosynthetic pathway); the known unknowns or "cryptic metabolites", which have never been isolated and characterized but whose existence can be inferred from biosynthetic pathways in sequenced genomes; and the unknown unknowns, or molecules whose existence has not been suspected through genomic or any other sort of analysis.

We do not know how many natural products exist, but the *Dictionary of Natural Products* lists 170,000 known structures, which effectively represents today's universe of known knowns. The biosynthetic pathways for most of these have not been definitively described, but progress on this front is extremely rapid, especially for molecules with significant biological activity produced by bacteria and fungi. The biosynthetic pathway encoding the genes that synthesize erythromycin's core, a molecule called erythronolide, condenses seven three-carbon fragments with additional possible transformations following each condensation, and because of the modular and repetitive nature of the biosynthetic chemistry, the gene cluster is also large (~50 kb) and highly repetitive (Figure 1b).[1,2] Biosynthetic gene clusters like the one making erythromycin are relatively easy to identify in sequenced genomes, and their identification has been automated with programs like antiSMASH.[3] The genes for many biosynthetic pathways can be manipulated relatively easily to create what have been called unnatural natural products.[4,5] The erythromycin pathway is arguably the world's best studied and most manipulated natural product biosynthetic pathway, and the *Dictionary of Natural Products* describes 34 biosynthetic relatives, some naturally occurring and some laboratory generated.

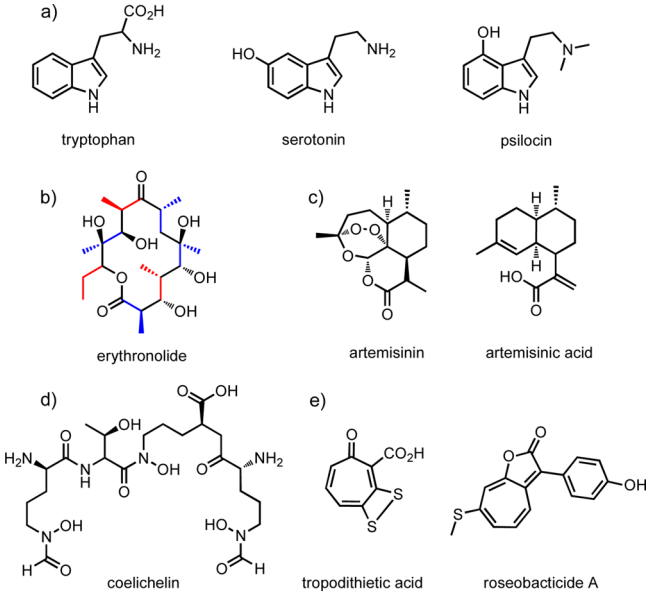

Figure 1. (a) Three molecules with similar structures and largely shared biosynthetic pathways, only one of which is a natural product. (b) Erythronolide A, the core of the antibiotic erythromycin, which is made by repetitive condensation of three carbon units as illustrated by the blue and red color scheme. (c) Artemisinin, the potent antimalarial drug, can be chemically synthesized from artemisinic acid, which can be made on a ton scale from a re-engineered biosynthetic pathway. (d) Coelichelin, a cryptic metabolite from *S. coelicolor*, that was identified bioinformatically and discovered with differential metabolomics. (e) Tropodithietic acid and roseobacticide A, a structurally unusual antibiotic and an algaecide, produced by *P. gallaeciensis*.

KNOWN KNOWNS

One important way synthetic biology connects with natural products of the known known variety is increasing the production of high-value materials with re-engineered pathways. The production of artemisinic acid, a key intermediate in the production of the important antimalarial drug artemisinin,

Special Issue: Natural Products

Received: March 10, 2013
Published: January 7, 2014

   dx.doi.org/10.1021/sb400025p | *ACS Synth. Biol.* 2014, 3, 745–747

represents a milestone achievement in this area (Figure 1c). For the past few thousand years artemisinin, an ingredient in traditional Chinese medicine, was obtained from the sweet wormwood plant (*Artemisia annua*), and today China, Vietnam, and East Africa are the major producers. The vagaries of agricultural production coupled with the sensitivity of malaria therapeutics to cost of goods considerations led Jay Keasling and co-workers to apply synthetic biology principles to artemisinin production.[6,7] The biosynthetic pathway leading to artemisinic acid, which can be converted to artemisinin in a straightforward if technically challenging synthetic step, was re-engineered in both a bacterial (*Eschericia coli*) and a yeast (*Saccharomyces cerevisiae*) host. Significantly, the approach differed from traditional heterologous expression by using hosts belonging to different taxonomic kingdoms from the original plant producer, and all of the biosynthetic steps in the pathway were recreated using genes from a variety of organisms. The resulting pathway differed from the known pathway in significant features. For example, the *E. coli* pathway did not use the bacteria's native deoxyxylulose-5-phosphate (DXP) mevalonate biosynthetic pathway but instead used the mevalonate pathway from yeast, and the enzymes were linked into a polyprotein rather than left as free-standing entities. The authors reported their initial success in 2006, and after a prolonged development period, Sanofi-Aventis just announced that it had produced 39 metric *tons* of artemisinic acid.[8]

KNOWN UNKNOWNS

Saccharopolyspora erythraea, the erythromycin producer, provides a convenient example of the known unknowns or cryptic metabolites category of natural products. The complete genome of *S. erythraea* became available in 2007 and contained some 17 other pathways that were likely to produce small molecules.[9] These pathways were easily identified because of their size and repetitive nature. The ratio of known to cryptic metabolites from *S. erythraea* has been mirrored in virtually all other bacterial genomes, and cryptic metabolites represent both a significant opportunity and a vexing challenge for natural products research. The opportunity for finding molecules with the potential utility of erythromycin in easily cultivated producers is tantalizing, but developing a broadly applicable approach to producing and identifying them has been quite difficult. Greg Challis and his co-workers pioneered one of the most successful and widely used approaches in their studies of *Streptomyces coelicolor*, the first fully sequenced member of the bacterial group that has produced the majority of useful therapeutic agents. The researchers first identified a cluster for a cryptic metabolite, named it coelichelin, and proposed a structure that could guide the search.[10] They then genetically disabled the candidate pathway and compared the suites of small molecules, the metabolomes, of wild type and mutant to identify a molecule present in wild type but absent in the mutant (Figure 1d).[11] While this *ad hoc* approach of generating targeted mutants and comparing metabolomes has been successful in many other projects, a universally applicable, high-throughput approach would greatly increase discovery rates. One tempting strategy, moving the pathways for cryptic metabolites to alternative hosts, has not been widely successful. Synthetic biology might provide a more fruitful approach by 'refactoring' the biosynthetic pathways of cryptic metabolites. In refactoring, which represents an even more thorough bottom-up re-engineering than the initial artemisinin pathways developed by the Keasling team, all noncoding DNA, regulatory proteins, and nonessential genes are deleted from the pathway.[12] The remaining genes are recoded to generate a DNA sequence that diverges as much as possible from that of the native pathway. These new genes are organized into operons and placed under the control of synthetic promoters, terminators, and ribosome binding sites to create a pathway with completely defined genetic parts that shares minimal DNA sequence similarity to the native pathway. Refactored pathways have been created for known knowns, which represent an important proof of principle, but success with known unknowns has not yet been reported.

UNKNOWN UNKNOWNS

With known knowns like artemisinin, synthetic biology supports natural products by improving supply; with known unknowns like cryptic metabolites, synthetic biology will likely support natural products by enabling discovery; and with unknown unknowns, natural products can inform synthetic biology. The discovery of unexpected natural products from new sources with new phenotypic screens will provide new tools for synthetic biology. As a modest example, consider tropodithietic acid and roseobacticide A, metabolites from marine proteobacteria that play key roles in a variable symbiosis between a marine microalga (*Emiliania huxleyi*) and the bacterial producer (*Phaeobacter gallaeciensis*) (Figure 1e).[13,14] Tropodithietic acid functions as an antibiotic for the symbiotic pair during the mutualist phase of the association, and roseobacticide A functions as an algal toxin during the antagonistic phase of the association. Both represent strikingly different chemotypes with largely unknown gene clusters and completely unknown biosynthetic pathways. However, when their biosynthetic pathways are known, they will provide powerful search tools for additional examples of these interesting if unfamiliar structures along with new tools, their biosynthetic enzymes, for the synthetic biology toolbox.

In the ways briefly outlined in this essay, one of the most venerable fields of chemistry, natural products, and one of the most exciting new fields, synthetic biology, can intersect to the mutual benefit of both.

AUTHOR INFORMATION

Corresponding Author

*Ph: 1-617-432-2845. Fax: 1-617-432-6424. E-mail: jon_clardy@hms.harvard.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The preparation of this article was supported by National Institutes of Health grants GM086258 (J.C.) and 1K99 GM098299 (M.R.S.).

REFERENCES

(1) Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. A., and Cane, D. E. (2007) Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annu. Rev. Biochem. 76*, 195−221.

(2) Cane, D. E. (2010) Programming of erythromycin biosynthesis by a modular polyketide synthase. *J. Biol. Chem. 285*, 27517−23.

(3) Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., and Breitling, R. (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res. 39*, W339−46.

(4) McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., and Ashley, G. (1999) Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. U.S.A. 96*, 1846−51.

(5) Baltz, R. H. (2006) Molecular engineering approaches to peptide, polyketide and other antibiotics. *Nat. Biotechnol. 24*, 1533−40.

(6) Keasling, J. D. (2012) Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng. 14*, 189−95.
(7) Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R., and Keasling, J. D. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature 440*, 940−3.
(8) Peplow, M. (2013) Malaria drug made in yeast causes market ferment. *Nature 494*, 160−1.
(9) Oliynyk, M., Samborskyy, M., Lester, J. B., Mironenko, T., Scott, N., Dickens, S., Haydock, S. F., and Leadlay, P. F. (2007) Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat. Biotechnol. 25*, 447−53.
(10) Challis, G. L., and Ravel, J. (2000) Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. *FEMS Microbiol. Lett. 187*, 111−4.
(11) Lautru, S., Deeth, R. J., Bailey, L. M., and Challis, G. L. (2005) Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nat. Chem. Biol. 1*, 265−9.
(12) Temme, K., Zhao, D., and Voigt, C. A. (2012) Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proc. Natl. Acad. Sci. U.S.A. 109*, 7085−90.
(13) Thiel, V., Brinkhoff, T., Dickschat, J. S., Wickel, S., Grunenberg, J., Wagner-Döbler, I., Simon, M., and Schulz, S. (2010) Identification and biosynthesis of tropone derivatives and sulfur volatiles produced by bacteria of the marine Roseobacter clade. *Org. Biomol. Chem. 8*, 234−246.
(14) Seyedsayamdost, M. R., Case, R. J., Kolter, R., and Clardy, J. (2011) The Jekyll-and-Hyde chemistry of *Phaeobacter gallaeciensis*. *Nat. Chem. 3*, 331−5.