Microbial genome mining answers longstanding biosynthetic questions

Jason M. Crawford[a] **and Jon Clardy**[b,1]

[a]*Department of Chemistry, Chemical Biology Institute, Yale University, New Haven, CT 06520; and* [b]*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115*

At the beginning of the genomic revolution, genomes were frequently described as the instruction manuals for an organism, but as the revolution progressed, many of these genomic instruction manuals resembled the frustratingly opaque descriptions that accompany so many new household gadgets, laboratory devices, and computer software. An important exception to this opacity has been in microbial natural product biosynthesis (1–7), and in PNAS, Davison et al. (8) provide a striking example of how the biosynthetic genes for a biologically important class of small molecules were spotted and analyzed to resolve a 70-y-old puzzle.

"Natural products" is the collective term for the wildly diverse families of small molecules produced by genetically encoded pathways. The natural products from fungi and bacteria have repeatedly transformed our understanding of biological systems and represent a substantial fraction of our current pharmaceuticals, especially those used as anticancer, antibiotic, or immunomodulatory agents (9). The biosynthetic pathways that give rise to natural products begin with the same small-molecule building blocks used by all of life, but they are joined together and modified into idiosyncratic molecules that differ dramatically from their universal forebears. Natural product biosynthetic pathways involve multiple genes, but in microbes the biosynthetic genes are most often clustered with regulatory and resistance genes on a contiguous stretch of the genome. With the right chemical, microbiological, or genetic tools, a natural product gene cluster can be connected to its cognate small-molecule product(s). In this way, natural product biosynthesis provides relatively easily studied multigenic phenotypes—the small-molecule products—that can be analyzed in a particularly informative way, because the contribution each gene makes to the final molecule can be determined. As microbial genomes become increasingly available, they are being mined both to discover new molecules with biomedical relevance and to find the answers to longstanding questions about how previously identified small molecules are produced.

The fungus *Talaromyces stipitatus* produces stipitatic acid (**6**), one of the founding members of an unusual class of fungal natural products characterized by a seven-membered ring. Most natural products—think of steroids like cholesterol, testosterone, and estradiol or aromatic amino acids like phenylalanine, tyrosine, and tryptophan—contain five- and six-membered rings, and seven-membered rings are relatively rare. They are also not usually found in synthetic medicinal agents. The seven-membered ring of stipitatic acid, called tropolone, is a key structural element found in a variety of mycotoxins, antibacterials, antifungals, and antimalarials (10). Until the report of Davison et al. (8), the enzymes involved in its construction had evaded detection for more than 70 y.

Davison et al. executed with precision two fundamental approaches to probing a small-molecule biosynthetic pathway. The first was a "top-down" biosynthetic approach, in which gene deletions are constructed in the producing organism, and protein function is inferred through loss of wild-type small-molecule production and buildup of pathway intermediates. The second was a "bottom-up" biosynthetic approach, in which an individual gene or set of genes is expressed in an alternative organism not known to produce the target molecule, and protein function is inferred through small-molecule product reconstitution. The group previously identified a nonreducing polyketide synthase (NR-PKS) in the related Ascomycota *Acremonium strictum* required for the synthesis of the aromatic polyketide 3-methylorcinaldehye (**1**) (11), a known precursor in tropolone biosynthesis (12), so the present study began with a search for related genes in the *T. stipitatus* genome. NR-PKSs are responsible for producing the diverse class of fungal aromatic polyketides (13, 14). The multidomain 3-methylorcinaldehyde NR-PKS contains a starter-unit-ACP-transacylase (SAT in Fig. 1) (ACP, acyl-carrier protein), a ketosynthase (KS), a malonyl-CoA:ACP transacyalse (MAT), a product template (PT) domain, tandem ACPs, a C-methyltransferase (*C*Met), and a reductive (R) domain that assembles four ketide (acetate-derived) units from acetyl/malonyl-CoA, methylates the growing chain once using S-adenosylmethionine, catalyzes regiospecific cyclization and aromatization, and releases the enzyme-bound product through a reductive release mechanism (Fig. 1) (15). Top-down deletion of the *T. stipitatus* PKS destroyed production of 3-methylorcinaldehye (**1**) and consequently shut down downstream tropolone biosynthesis. Linking gene deletion with loss of production of 3-methylorcinaldehyde (**1**) confirmed the identity of the cluster. Likewise, bottom-up reconstitution using an intronless version of the NR-PKS in the heterologous host *Aspergillus oryzae* led to production of **1**. These initial results provided the basis for launching a detailed investigation of the biosynthetic proteins encoded in the surrounding tropolone gene cluster.

Oxidative processing represents a common biosynthetic strategy for generating structurally complex molecules from relatively simple starting materials, and the route from 3-methylorcinaldehyde (**1**) to stipitatic acid (**6**) requires a sequence of undescribed oxidative modifications and oxidative rearrangements. Continuing with the top-down and bottom-up themes, Davison et al. (8) targeted three oxidative enzymes in the gene cluster, a flavin adenine dinucleotide (FAD)-dependent monooxygenase, a cytochrome P450 monooxygenase, and a nonheme iron oxygenase. These genes were individually knocked out in *T. stipitatus* to monitor the effects in product distribution and were individually expressed as His-tagged recombinant proteins in *Escherichia coli* to more finely query the function of purified enzymes, the ultimate experimental demonstration of the bottom-up biosynthetic approach.

Deletion of the FAD-dependent monooxygenase (*tsl1*/*tropB*) led to accumulation of 3-methylorcinaldehyde (**1**) and identification of side product **7** that underwent 5-hydroxylation by an unknown oxidase in the producing organism (Fig. 1). Accumulation of **1** suggested that TropB was an early enzyme in the oxidative processing sequence. Indeed, an in vitro reaction with the purified FAD-dependent trimeric TropB protein with 3-methylorcinaldehyde (**1**) and NADPH catalyzed oxygenation and dearomatization of **1**

Author contributions: J.M.C. and J.C. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 7642.

[1]To whom correspondence should be addressed. E-mail: jon_clardy@hms.harvard.edu.

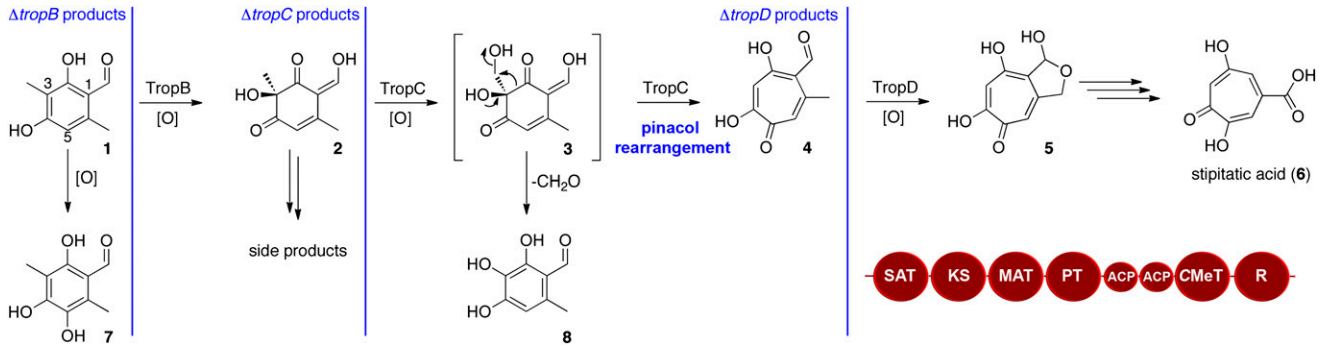

Fig. 1. Biosynthetic highlights for stipitatic acid. *Inset*: The arrangement of biosynthetic domains that lead to the production of 3-methylorcinaldehyde (**1**). Key oxidizing enzymes, their reactions, and important by-products: three oxidative enzymes, TropB, TropC, and TropD, convert 3-methylorcinaldehyde (**1**) to tropolone **5**. TropC has two functions, the oxidation of a methyl to a hydroxymethyl and the pinacol rearrangement of **3** to **4**.

to yield the 3-hydroxylation product **2**, demonstrating that TropB initiates the oxidative cascade. The nonheme iron oxygenase (*tsr5*/*tropC*) deletion accumulated the TropB product **2** in addition to non–wild-type side reactions that occurred in the presence of this elevated product, suggesting that TropC acted next in the sequence on-path to the tropolone nucleus **4**. Reconstitution of the purified recombinant monomeric nonheme iron oxygenase TropC with **2**, FeII, α-ketoglutarate, and ascorbate resulted in the synthesis of tropolone **4** and a minor by-product **8**, defining the critical enzyme involved in catalyzing the oxidative ring expansion to the seven-membered tropolone nucleus. The final oxidative cytochrome P450 enzyme TropD resisted experimental measures to obtain soluble, catalytically competent protein from *E. coli*, but deletion of *tropD* in the fungal host led to accumulation of the TropC products **4** and **8**. Again, these results supported that TropD functions next in the oxidative sequence, and the authors hypothesized that TropD hydroxylates **4** to **5** on path to stipitatic acid (**6**).

The oxidative ring expansion in tropolone biosynthesis has intrigued chemists for many decades, and the mechanisms involved in its formation were largely probed through isotopic labeling studies (10). Proposed monooxygenase and dioxygenase mechanisms were originally envisioned. O'Sullivan and Schwab (16) conducted an $^{18}O_2$ labeling experiment to distinguish between the two routes, and after measuring a single ^{18}O incorporation in stipitatic acid (**6**), concluded that the favored mechanism proceeded through a hydroxymethyl intermediate that underwent a pinacol rearrangement to generate the expanded tropolone ring (Fig. 1). Isotopic labeling studies of a given product provide a "black box" summary of the chemical transformations that occur from a simple labeled substrate to a fully advanced product, but identification of specific enzymes involved in often complex transformations remains enigmatic. Although a hydroxymethyl intermediate was not detected in the present study, Davison et al. (8) illuminate TropC as the specific oxidative enzyme that most plausibly proceeds through intermediate **3** as a prelude to the TropC-mediated pinacol rearrangement that produces the tropolone product **4**. The authors propose that identification of the minor by-product **8** in this enzymatic reaction further supports intermediacy of **3** via incomplete enzymatic conversion to the seven-membered ring.

Now that the functions of the oxidative enzymes in tropolone biosynthesis have been ascribed, these proteins represent powerful mining tools to uncover related tropolone-derived bioactive molecules encoded in the rapidly growing panoply of publicly available genomic information. Indeed, the authors already identified related genes in 10 other organisms. Furthermore, because tropolone represents a common nucleus in multiple pathways, the authors reasonably speculate that similar reactions will be at play in the biosynthesis of other important fungal compounds, such as xenovulenes, epolones, ditropolones, citrinins, sclerotiorins, and asperfuranones. Although tropolone-derived natural products represent the striking example discussed here, combing microbial genome data with a keen biosynthetic eye will undoubtedly lead to other new biosynthetic enzymes and new small-molecule structural classes or features, and these small molecules could in turn lead to a deeper understanding of the connection between naturally produced small molecules and the biological processes they control.

1. Nett M, Ikeda H, Moore BS (2009) Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep* 26:1362–1384.
2. Wenzel SC, Müller R (2009) The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Nat Prod Rep* 26:1385–1407.
3. Kalaitzis JA, Lauro FM, Neilan BA (2009) Mining cyanobacterial genomes for genes encoding complex biosynthetic pathways. *Nat Prod Rep* 26:1447–1465.
4. Gross H, Loper JE (2009) Genomics of secondary metabolite production by *Pseudomonas spp*. *Nat Prod Rep* 26:1408–1446.
5. Brady SF, Simmons L, Kim JH, Schmidt EW (2009) Metagenomic approaches to natural products from free-living and symbiotic organisms. *Nat Prod Rep* 26:1488–1503.
6. Sanchez JF, Somoza AD, Keller NP, Wang CC (2012) Advances in *Aspergillus* secondary metabolite research in the post-genomic era. *Nat Prod Rep* 29:351–371.
7. Winter JM, Behnken S, Hertweck C (2011) Genomics-inspired discovery of natural products. *Curr Opin Chem Biol* 15:22–31.
8. Davison J, et al. (2012) Genetic, molecular, and biochemical basis of fungal tropolone biosynthesis. *Proc Natl Acad Sci USA* 109:7642–7647.
9. Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod* 75:311–335.
10. Bentley R (2008) A fresh look at natural tropolonoids. *Nat Prod Rep* 25:118–138.
11. Bailey AM, et al. (2007) Characterisation of 3-methylorcinaldehyde synthase (MOS) in *Acremonium strictum*: First observation of a reductive release mechanism during polyketide biosynthesis. *Chem Commun (Camb)* 39: 4053–4055.
12. Bryant RW, Jr., Light RJ (1974) Stipitatonic acid biosynthesis. Incorporation of (formyl-^{14}C)-3-methylorcylaldehyde and (^{14}C)stipitaldehydic acid, a new tropolone metabolite. *Biochemistry* 13:1516–1522.
13. Zhou H, Li Y, Tang Y (2010) Cyclization of aromatic polyketides from bacteria and fungi. *Nat Prod Rep* 27:839–868.
14. Crawford JM, Townsend CA (2010) New insights into the formation of fungal aromatic polyketides. *Nat Rev Microbiol* 8:879–889.
15. Fisch KM, et al. (2010) Catalytic role of the C-terminal domains of a fungal non-reducing polyketide synthase. *Chem Commun (Camb)* 46:5331–5333.
16. O'Sullivan MC, Schwab JM (1995) Verification of the mechanism of oxidative ring expansion in the biosynthesis of stipitatic acid by *Talaromyces stipitatus*. *Bioorg Chem* 23:131–143.