

UCLA

UCLA Previously Published Works

Title

Deciphering chemical logic of fungal natural product biosynthesis through heterologous expression and genome mining.

Permalink

<https://escholarship.org/uc/item/2d89695z>

Journal

Natural Product Reports: a journal of current development in bioorganic chemistry, 40(1)

Authors

Chiang, Chen-Yu

Ohashi, Masao

Tang, Yi

Publication Date

2023-01-25

DOI

10.1039/d2np00050d

Peer reviewed



Published in final edited form as:

Nat Prod Rep. ; 40(1): 89–127. doi:10.1039/d2np00050d.

Deciphering Chemical Logic of Fungal Natural Product Biosynthesis through Heterologous Expression and Genome Mining

Chen-Yu Chiang¹, Masao Ohashi^{1,*}, Yi Tang^{1,2,*}

¹Dept. of Chemical and Biomolecular Engineering, 5531 Boelter Hall, 420 Westwood Plaza, Los Angeles, CA 90095, USA

²Dept. of Chemistry and Biochemistry, 5531 Boelter Hall, 420 Westwood Plaza, Los Angeles, CA 90095, USA

Abstract

Heterologous expression of natural product biosynthetic gene clusters (BGCs) has become a widely used tool for genome mining of cryptic pathways, bottom-up investigation of biosynthetic enzymes, and engineered biosynthesis of new natural product variants. In the field of fungal natural products, heterologous expression of a complete pathway was first demonstrated in the biosynthesis of tenellin in *Aspergillus oryzae* in 2010. Since then, advances in genome sequencing, DNA synthesis, synthetic biology, etc. have led to mining, assignment, and characterization of many fungal BGCs using various heterologous hosts. In this review, we will highlight key examples in the last decade in integrating heterologous expression into genome mining and biosynthetic investigations. The review will cover the choice of heterologous hosts, prioritization of BGCs for structural novelty, and how shunt products from heterologous expression can reveal important insights into the chemical logic of biosynthesis. The review is not meant to be exhaustive but is rather a collection of examples from researchers in the field, including ours, that demonstrate the usefulness and pitfalls of heterologous biosynthesis in fungal natural product discovery.

1. Introduction

Genome mining, a term first mentioned in the context of natural product in 2005,¹ has brought a renaissance to the research fields of natural product discovery, biosynthesis and chemical biology. More recently, the marriage of biosynthesis and synthetic biology has further broadened interests in natural products. Advances in next-generation DNA sequencing, DNA synthesis and gene editing technologies have rapidly enhanced our abilities to identify, construct and characterize biosynthetic gene clusters (BGCs). Among the different approaches to mine new natural products or rediscover biological activities of known compounds, heterologous expression of BGCs in model organisms is a key strategy.^{2,3} Heterologous expression of BGCs, coupled with host engineering, has been successfully demonstrated for nearly all major families of natural products from bacterial,

* yitang@ucla.edu, gph422001@ucla.edu.

fungal, plant and animal pathways.⁴⁻⁷ Our lab has been involved in the genome mining of natural products from filamentous fungi in the last decade. Genome sequencing of Ascomycota, such as *Penicillium* and *Aspergillus* species, have revealed a significant amount of BGCs are unexplored and have no associated metabolites.^{8,9} It is estimated that over 97% of fungal BGCs are not associated with known natural products, a number that is in line with that estimated for bacterial BGCs.^{10,11}

In this review, we highlight a number of examples from ours and other researchers in genome mining of fungal BGCs. In section 2, we describe the general workflow of BGC refactoring and expression in different heterologous hosts. Several commonly used hosts, such as *Aspergillus nidulans*, *Aspergillus oryzae*, and *Saccharomyces cerevisiae* are discussed with representative examples. Considerations to increase successes of reconstitution are discussed. In section 3, we focus on strategies to expand natural product chemical space through mining of BGCs with unique features. In section 4, we recount several case studies in which intermediates and shunt products from heterologous reconstitution led to new discoveries of biosynthetic logic. Collectively, this review illustrates the successes and failures of genome mining in fungi, and underscores the vast unexplored biosynthetic potential in this kingdom of eukaryotic organisms.

2. Overview of Heterologous Expression of Fungal Natural Products

2.1. Why heterologous expression?

While the focus of this review is on the use of heterologous hosts for fungal BGC expression, the importance of working with native producing strains cannot be overstated. If the original producing host can be sourced from the many different collections around the world, working directly in a genome-sequenced strain has unmatched advantages. First, the strain can be grown under a plethora of media conditions to produce different metabolic profiles. This is the “one strain many compounds” (OSMAC) strategy that can activate a subset of BGCs.¹²⁻¹⁴ Second, the compounds produced in a native host are highly likely to be the *final* natural product of the pathway; and may be produced at sufficient titers for full structural characterization. This is especially important in cases where the BGC is too large to be refactored using heterologous host/vector systems, or has ambiguous definitions of cluster boundaries.¹⁵ Lastly, if the fungal strain can be genetically modified, many tools can be directly applied to the native host including BGC-specific transcriptional factor (TF) overexpression,¹⁶⁻¹⁹ global epigenetic modifications,^{20,21} CRISPR-Cas9 mediated pathway activation,²² and top-down genetic inactivation to assign functions to individual pathway genes (Figure 1).²³

Clearly, if the native strain cannot be sourced or is genetically intractable, transplanting the BGC of interest into a chassis organism for heterologous expression is the only alternative. Heterologous hosts are genetically well-characterized with robust molecular biology tools. An ideal heterologous host is also fast growing and has a minimal metabolic background. For bacterial BGCs, *E. coli* and different *Streptomyces* organisms are used frequently.²⁴⁻²⁷ For plant BGCs, *Saccharomyces cerevisiae* (Baker's yeast, also referred to as yeast throughout the review) is the go-to microbial chassis,²⁸ while the tobacco plant *Nicotiana benthamiana* is the most used plant heterologous host.²⁹ For fungi, a number of

strains are available, including *Saccharomyces cerevisiae* and well-characterized *Aspergillus* sp. that are discussed in section 2.2. Several other fungal species have also been reported to be useful heterologous hosts.^{30,31}

Even when the native fungal host can be genetically manipulated, heterologous expression offers several enabling advantages in analysis of BGCs. First, with a reduced metabolic background and rewired precursor fluxes dedicated to heterologously expressed pathways, the titers of the target metabolite may be engineered to be significantly higher than from the native host. However, metabolite titers in a heterologous host can be highly variable and depend on the BGC of interest. Titters from less than 1 mg/L to well over 1 g/L have been reported under similar growth conditions for different BGCs. The exact reasons for such variation are not well-understood; and empirical methods are used to optimize the titers. For the purpose of this review, titers are generally not discussed, and readers should refer to the primary literature for such information. Second, many BGCs in native hosts cannot be transcriptionally activated despite the myriad of tools mentioned above. Through complete refactoring of BGCs for heterologous expression, such cryptic regulation can be bypassed. Third, if the native strain is difficult to obtain due to logistical reasons, reconstitution of homologous BGC from accessible organisms can overcome such hurdles. For example, in our effort to study sambutoxin (**1**) biosynthesis, the producing strain *Fusarium sambucinum* was not obtainable.³² Comparison of candidate BGCs revealed the wide-spread presence of the candidate BGC in other species, including *Fusarium oxysporum* of which the genomic DNA can be obtained. Heterologous reconstitution of the six genes in the BGC in *A. nidulans* confirmed the cluster indeed produced sambutoxin. Similarly, the biosynthesis of the potent oxidative phosphorylation inhibitor ilicicolin H (**2**) was reconstituted in *A. nidulans* by expressing a homologous BGC from *Penicillium variable*, which was not known to be a producer prior to the study.³³ Lastly, with a modular refactoring approach, mix-and-matching of pathway genes between target BGC and homologous BGCs can be readily performed in heterologous hosts. This can overcome expression or pathway flux bottleneck attributed to one particular enzyme from a BGC, as well as to generate structural variants of natural products.^{34,35}

Heterologous reconstitution enables the gene-by-gene, bottom-up approach to examine individual steps of biosynthesis and allows functional association between BGCs and natural products in an unambiguous way (Figure 1). While the genetic inactivation-based, top-down approach in native hosts can be highly informative, several examples of misinterpretation of enzyme function have been noted in the literature due to pleiotropic effects of gene inactivation. We note two examples here, in which subsequent heterologous reconstitution studies led to revision of initial BGC assignments. In the first example, Chankhamjon et. al. identified AoiQ as the halogenase responsible for biosynthesis of 8-methyldichlorodioporthin (**3**) from *Aspergillus oryzae* RIB40,³⁶ which is a coumarin-containing polyketide with a *gem*-dichlorinated *sp*³ carbon α to a secondary alcohol. Based on genetic knockout in the native host, the authors concluded an unclustered, nonreducing polyketide synthase (NRPKS) AoiG is involved in biosynthesis of the unchlorinated substrate. This led to the speculation that AoiQ, a flavin-dependent halogenase, can chlorinate an unactivated *sp*³ carbon via an unprecedented radical mechanism. Subsequent

heterologous expression work, however, showed a different NRPKS DiaA, which is clustered with AoiQ, is involved in synthesizing the coumarin core.³⁷ DiaA synthesizes a 1,3-diketone containing substrate that can be halogenated by AoiQ with canonical flavin-dependent halogenation mechanism via an enolate intermediate. Nonenzymatic deacetylation and short-chain dehydrogenase/reductase (SDR)-catalyzed ketoreduction afforded dichlorodiaporthin **3** (Figure 2A).

The second example of misinterpretation of gene deletion results is that of thermolide biosynthesis.³⁸ Thermolides are a class of nematocidal natural products isolated from the thermophilic fungus *Talaromyces thermophilus* NRRL2155, featuring an uncommon macrolactone linkage between the polyketide and nonribosomal peptide portions. Inactivation of the polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS) TalA in the native producing organism abolished the production of thermolides. This led to the putative assignment of TalA as the scaffold-generating enzyme, although the predicted functions of remaining enzymes in the *tal* BGC were not consistent with thermolide structural features. The Zou lab subsequently performed heterologous expression of all candidate BGCs in *A. nidulans* to functionally characterize the PKS-NRPSs. They demonstrated that the *tal* BGC was in fact responsible for biosynthesis of analogs of fusarin C. The *bona fide* thermolide BGC encodes a PKS (ThmA) and an NRPS (ThmB) as two separate proteins, with a terminal C domain (C_T) in the NRPS enzyme catalyzing the macrolactonization step (Figure 2B). Coexpression of ThmA and ThmB led to the production of the macrolactone **4**. The earlier gene inactivation of *talA* may have caused pleiotropic silencing of the thermolides BGC.

2.2. Host/vector systems for heterologous expression of fungal BGCs

To date, several well-developed organisms have been used as microbial chassis for heterologous expression of fungal BGCs. Useful heterologous hosts share many advantageous features: fast growth under laboratory culturing conditions; genetical tractability; and not posing biosafety hazards when grown in large volumes. Additional considerations in choosing the model organism include i) the number of heterologous genes that can be expressed from episomal vectors or chromosomal integrations; ii) crosstalk with endogenous metabolism (both primary and secondary) or detoxification pathways; iii) endogenous secondary metabolite levels that can interfere with compound identification or purification; and iv) most importantly, abilities to correctly splice introns during mRNA maturation and perform essential post-translational modifications (PTMs) of foreign biosynthetic enzymes.

2.2.1. *Saccharomyces cerevisiae*—Although a number of fungal hosts have been designed and used successfully as heterologous hosts,^{39–42} three hosts have emerged to be the workhorse organisms. Baker's yeast, *Saccharomyces cerevisiae*, is widely used as a model eukaryote for research in biology, as well as food fermentation and industrial synthetic biology applications. It has been adopted for the reconstitution of biosynthetic pathways because of several attractive traits, including fast growth rate, generally regarded as safe (GRAS) status, accessible genetic tools, and very few endogenous secondary metabolites produced. Metabolic engineering of yeast is also highly advanced, producing

nearly every major type of natural products at high titers (g/L), including polyketides, terpenes, and alkaloids.^{43–48} One of the earlier engineering efforts to make yeast a suitable host for fungal polyketide and nonribosomal peptide biosynthesis was the integration of the *A. nidulans* phosphopantetheinyl (pPant) transferase *npgA* gene into the genome of BJ5464 to give BJ5464-NpgA^{49,50}. NpgA phosphopantetheinylates the active site serines in thiolation domains (ACP and PCP) in PKS and NPRS enzymes, which is essential for the thioester chemistry operated by these machineries. BJ5464-NpgA is analogous to the *E. coli* strain BAP1,⁵¹ in which the bacterial pPant transferase Sfp is expressed chromosomally from an engineered BL21(DE3) strain. An additional feature of BJ5464-NpgA, which is vital for its use in expression of large fungal biosynthetic enzymes, is the knockout of vacuole proteases PEP4 and PRB1. With these genetic changes, BJ5464-NpgA was shown to be suitable for functional expression of fungal megasynthases, including PKS (LovB, 330 kDa),⁵² NRPS (TqaA, 450 kDa)⁵³ and PKS-NRPS (AspA, 493 kDa).⁵⁴ Our lab further modified the strain to give RC01,⁵⁵ which expresses a chromosomal integrated copy of *Aspergillus terreus* cytochrome P450 reductase (AtCPR). This redox partner enzyme is required for fungal P450 catalytic function as it provides two electrons, one at a time, to the P450 heme center during the catalytic cycle.⁵⁶ With this integration, fungal pathways involving multiple P450s have been reconstituted or mined.^{55,57}

High-throughput and scalable genome mining of fungal BGCs using yeast as the heterologous host was reported using the HEx platform.⁵⁸ In this study, the host JHY702 was a heavily modified strain to combine the above features, including integration of *npgA* and *atCPR*, as well as modifications of genes involved in sporulation, respiratory growth, and protein expression, to improve strain robustness. Furthermore, the study examined the use of different ADH2-like promoters, which are inducible promoters activated upon glucose depletion and ethanol accumulation, from various yeast species. The sequentially divergent but functionally equivalent promoters allowed rapid pathway refactoring through yeast homologous recombination. Using HEx platform, 22 out of 41 tested BGCs from diverse fungal species produced detectable compounds, including polyketides, peptides and terpenes.

Despite the ease of use of yeast as a model host, several deficiencies have severely limited its application in genome mining. The most glaring is its inability to splice fungal introns. Although yeast has a spliceosome machinery,^{59,60} it is not able to recognize nor process fungal introns. This deficiency requires fungal genes with introns to be manually spliced to give uninterrupted open reading frame (ORFs) for yeast expression. As a result, accurate prediction of start codons and introns are indispensable to successful reconstitution. Often times, this would require isolation of mRNA from the native host, followed by RT-PCR and sequencing; or by first expressing the gene in a fungal heterologous host followed by mRNA isolation if the BGC of interest is silent. The limitation of intron splicing was attributed as a major reason for the failed reconstitution of several BGCs using the Hex platform.⁵⁸ In addition, we and others have noted foreign P450 expression in yeast, especially multiple P450s, can be highly unpredictable and problematic.⁶¹ As a result of these limitations, the engineered yeast strains are limited to reconstitution of core biosynthetic enzymes or shorter pathways, while leaving genome mining of more complex BGCs to the filamentous fungi discussed below.

2.2.2. *Aspergillus oryzae*—In terms of filamentous fungi, *Aspergillus oryzae* is one of the most commonly used heterologous hosts for BGC reconstitution due to many of the advantages previously discussed. Two modified strains, M-2-3 and NSAR1, are used by numerous labs for genome mining. *A. oryzae* M-2-3 is an auxotroph for arginine,⁶² and pTAex3, which encodes the *argB* gene from *A. nidulans*, is used as the corresponding vector for transformation and selection. *A. oryzae* NSAR1 was engineered to harbor four auxotrophic markers (*niaD*⁻; *sC*⁻; *argB*; *adeA*⁻),⁶³ making it a more versatile strain for genetic manipulation and full pathway reconstitution. Two additional vectors, with either pyrithiamine (pPTRI) or glufosinate (pBAR) resistance markers, can be introduced into M-2-3 or NSAR1. In the case of NSAR1, a total of six vectors can be transformed at the same time.

Cox and coworkers reported the first successful reconstitution a complete fungal BGC, that of tenellin (**5**),⁶⁴ in *A. oryzae* M-2-3 using *argB* auxotroph as well as glufosinate and bleomycin resistant markers. Refactoring the cluster and placing the genes under control of inducible *amyB* promoters, the titer of **5** (243 mg/L) was five times higher than that of the native host.⁶⁴ In the same year, heterologous reconstitution using the same strain to study biosynthesis of fungal meroterpenoid pyripyropene A (**6**) was published by Abe and coworkers.⁶⁵ Five biosynthetic genes were individually cloned into pTAex3 and pPTRI. Coexpression of different combinations of gene with feeding of precursors led to the discovery of early biosynthetic steps to pyripyropene A. In subsequent years, the Abe group has used *A. oryzae* extensively for meroterpenoid biosynthesis reconstitution.⁶⁶ Stepwise reconstitution, followed by isolation and characterization of biosynthetic intermediates and shunt products enabled functional assignment of many meroterpenoid pathways. The Oikawa group has used *A. oryzae* to examine the biosynthesis of indole-diterpene natural products such as paxilline (**7**).⁶⁷ One feature of the *A. oryzae* system is that the plasmids designed for protein expression cannot be autonomously replicated, therefore requiring all genes to be integrated into the genome for expression. As a result, colony-to-colony gene expression variation is an issue since the integrated sites can vary in each transformant. To have more control over gene integration and expression levels, Liu et al. identified high expression loci in *A. oryzae*, and developed a CRISPR-Cas9-based technique for targeted chromosomal integration.⁶⁸ This new strategy successfully led to the reconstitution of erinacine Q (**8**) and its intermediates in higher titer (4.7 mg/L for **8**).

Although *Aspergillus* heterologous hosts can splice most fungal introns correctly, one must remain cautious that incorrect splicing, which is detrimental to reconstitution, can occur. One published example of incorrect intron processing in *A. oryzae* was reported by Song et al. during the expression of a *Magnaporthe oryzae* PKS-NRPS encoded by *ACE1*.⁶⁹ RT-PCR of *ACE1* expression revealed that one of three introns was not spliced correctly, resulting in no production of any metabolites from the *A. oryzae* transformant. Further heterologous expression attempts using intron-free DNA led to detection of new metabolites, confirming incomplete splicing was the obstacle. Intron splicing is also an issue when expressing Basidiomycota genes in *Aspergillus* strains, since genes from Basidiomycota can contain many introns. Thus, intron-free cDNAs are typically required for refactoring the corresponding pathways.⁷⁰ Recently, *A. oryzae* was examined as a suitable host

for expressing Basidiomycota diterpene genes directly using genomic DNA sequences.⁷¹ Nagamine et al. screened thirty terpene synthases from two Basidiomycota strains to show that *A. oryzae* can splice most of those genes correctly. In cases where partial splicing is observed, the errors were analyzed and were mostly predictable according to the authors.

2.2.3. Aspergillus nidulans—*Aspergillus nidulans* is another important model organism for filamentous fungi and it has been used as heterologous host for exploration of many BGCs in recent years. *A. nidulans* A1145 (*pyrG89*; *pyroA4*; *nkuA::argB*; *riboB2*) is the most widely used strain and has three auxotrophic markers, *pyrG* (uracil), *pyroA* (pyridoxine), and *riboB* (riboflavin).⁷² Compared to *A. oryzae*, the *A. nidulans* A1145 host can be used to express heterologous genes from either episomal vectors or chromosomal integrations. In the former case, vectors contain the plasmid replicator AMA1 can be maintained and replicated in the host under selection pressure. Constitutive and nutrient-inducible promoters involved in primary metabolism, such as *glaA*, *gpdA*, and *amyB*, are used for controlling expression of foreign genes. Yeast-fungi-*E. coli* shuttle vectors have been created for construction of plasmids using homologous recombination in yeast and amplification by *E. coli*.⁷³ As many as twelve genes have been expressed at the same time using the host/vector pair, enabling numerous genome mining and biosynthetic reconstitution studies. Recently, a platform using native promoters to express heterologous genes in *A. nidulans* was reported.⁷⁴ The approach used regulatory components of the highly expressed asperfuranone BGC (*afo*) in *A. nidulans*. Induction of AfoA, which is a pathway-specific transcriptional activator, can lead to high titer production of asperfuranone and intermediates (~900 mg/L).⁷⁵ Chiang et al. demonstrated that genes of interest can be placed under *afo* promoters, allowing for the strong and concerted transcription upon AfoA expression. As a proof of concept, the authors used this platform to reconstitute biosynthesis of citreoviridin, mutilin, and pleuromutilin in *A. nidulans* with improved titers (~150 mg/L for pleuromutilin). This concept is akin to that developed by Kakule et al. in using *Fusarium heterosporum* as a heterologous host.³⁹ The limitation of this refactoring strategy is that the maximum number of genes introduced in the heterologous host cannot exceed the number of genes in the original cluster.

One disadvantage of *A. nidulans* is its robust endogenous secondary metabolism. With well over fifty identified BGCs, the host can produce a significant background of compounds under laboratory cultivating conditions. Furthermore, upon introduction of the replicative-competent vectors, additional metabolic pathways can be activated which lead to emergence of new compounds that can complicate target BGC analysis. To alleviate this problem, multiple labs have engineered the host to abolish some of the most abundant endogenous metabolites. For example, our lab used CRISPR-Cas9 facilitated homologous recombination to delete *easA* and *stcA*, responsible for emericellamide (**9**) and sterigmatocystin (**10**) biosynthesis, respectively, to arrive at *A. nidulans* A1145 ST EM.⁷⁶ These knockouts enabled the analysis of the zaragozic acid (**11**) pathway, of which the LC-MS features of a key intermediate were masked by the presence of sterigmatocystin prior to *stcA* deletion. The Wang lab reported a highly engineered strain LO8030 with deletion of eight endogenous BGCs,⁷⁷ which facilitated the discovery of aspercryptin (**12**) as an additional endogenous metabolite in *A. nidulans* due to a cleaner metabolic profile. The Keller lab

developed *A. nidulans* RJW256, which is an auxotrophic strain of *pyrG89* and *pyroA4* with the deletion of sterigmatocystin gene cluster, for applications in working with fungal artificial chromosomes (FACs).^{78,79}

In addition to the three model hosts discussed above, a number of other fungal species have been evaluated and engineered to be a chassis for fungal BGC expression. These strains include *Aspergillus niger*,⁴⁰ *Penicillium rubens*,⁴² *Fusarium heterosporum*,³⁹ *Fusarium graminearum*,³⁰ *Trichoderma reesei*³¹ etc. While each of these strains offers some unique advantages (titer, splicing, etc) over the *Aspergillus* strains for specific BGCs, they are not as widely adopted for natural product genome mining applications.

3. Examples of genome mining guided by biosynthetic gene cluster features.

Most researchers performing natural product genome mining are interested in discovering one or more of the following: 1) new chemical structures that have not been previously reported; 2) natural products with new or more potent biological activities; and 3) new enzymes catalyzing synthetically challenging reactions. With an estimated more than one million BGCs from microbial genome sequences, a key step is prioritization of the clusters based on individual research objectives. The approach for finding new biological activity is certainly different from finding novel chemical scaffolds, although one would expect new activities should reside in new chemical space.

With a chemocentric focus for this review, we will not elaborate on genome mining for new biological activities, except a brief detour on the topic of self-resistance gene guided search. This has been extensively reviewed by us and others recently.^{80,81} Briefly, the premise of this approach is that for a producing host to survive the (potent) activity of a natural product, proteins or enzymes that can confer self-resistance must be coexpressed with the pathway enzymes in the BGC. In the case of a self-resistance enzyme (SRE), it is a homolog of the housekeeping target of the natural product, but sufficiently mutated to confer resistance and keep the producing host alive. The colocalization of SRE with biosynthetic enzymes in a BGC provides a predicative window of the biological activity of the molecule encoded in the BGC; and can be leveraged to query database of BGCs for the desired biological activity. This colocalizations phenomenon is widely observed in both bacteria and fungi, and has been used to find BGCs of compounds with known activities, such as fumagillin (**13**)⁸², fellutamide (**14**)⁸³ (Scheme 2A); or to assign functions to known natural product following identification of BGCs.⁸⁴ One application of SRE-guide genome mining is the successful identification of a natural product inhibitor for fungal and plant dihydroxyacid dehydratase (DHAD) involved in branched chain amino acid biosynthesis (Figure 3).⁵⁷ Using yeast as the heterologous host, a terpene biosynthetic pathway encoding one terpene cyclase, two P450s and a putative self-resistant DHAD was shown to produce the submicromolar inhibitor aspterric acid (**15**). **15** is a promising herbicide lead and represents the first reported natural product inhibitor of DHAD. Following reconstitution in *A. nidulans* and identification of a SRE, the previously known fungal natural product harzianic acid (**16**) was shown to be an inhibitor of acetolactate synthase (ALS),⁸⁵ another enzyme in the branched

chain amino acid pathway. **16** binds to ALS in a different mode compared to all synthetic inhibitors, which can explain how it can evade the widely found resistance mutations. Overall, while only a small fraction of BGCs contain SREs and some putative SREs turn out to be biosynthetic rather than self-resistant, this approach offers the promise to rapidly associate biological activity to BGC-driven genome mining.

To prioritize genome mining towards new natural product structures and enzymes, we can consider the classification officially coined by Biermann and Helfrich,⁸⁶ in which a quadrant system (2×2) relating natural product structures to BGCs is used (Figure 4). Here both the BGC and natural products can be classified as either known or unknown. A known BGC refers to one that can be predicted to make a certain compound class based on the presence of a canonical core biosynthetic enzyme, such as polyketide synthase (PKS), nonribosomal peptide synthetase (NRPS), terpene synthase/terpene cyclase (TS/TC), prenyltransferase (PT), etc. These core enzymes are the basis of bioinformatics algorithms that catalog BGCs from sequenced genomes.⁸⁷ An unknown BGC refers to one that has no identifiable core enzyme, and the compound type cannot be readily predicted. With regard to the natural product, known vs unknown simply refers to whether the compound has been identified and structurally characterized. One can place most of the genome mining efforts into one of the four quadrants. Once the BGC-natural product association has been confirmed using native or heterologous host, that pair is placed in the known (BGC)-known (metabolite) category (I). If the natural product structure does not readily suggest a biosynthetic origin and the cluster is unknown, then that pair is placed into quadrant of unknown (BGC)-known (metabolite) (III). These compounds can be exciting targets to pursue new biosynthetic chemistry, as represented by recent discoveries of BGCs for altemicidin (**17**),⁸⁸ fluopsin C (**18**),⁸⁹ and guanitoxin (**19**)⁹⁰ (Scheme 2B). The known (BGC)-unknown (metabolite) (II) is where most of the BGC-driven genome mining activities originate and are discussed in detail below. The last quadrant (IV), which represents the true biosynthetic dark matter, is the unknown (BGC)-unknown (metabolite) category, in which BGCs of unknown functions are predicted to produce new natural product structures. To focus on efforts in unknown-unknowns, researchers must deemphasize or deprioritize known core enzymes. While this represents the most challenging and nascent area of genome mining, it is likely that truly novel chemical structures and biological activities will arise from BGCs currently classified in this quadrant.

The known-unknowns (II in Figure 4) represent most of the ongoing genome mining efforts in both bacteria and fungi. Using the known core enzymes as queries, a catalog of BGCs can be rapidly generated from each genome. In fungi that are prolific producers of natural products, it is typical to find more than 50 BGCs per genome using PKS, NRPS, TS/TC as leads. Enzymes that are frequently found in natural product BGCs, such as prenyltransferases (PTs) and RiPPs maturation enzymes can also populate the bioinformatics query. Compared to bacteria, other families of natural products such as aminoglycosides, phosphonates, etc. are not typically produced by fungi. To prioritize BGCs that may produce new chemical structures, several search criteria have been employed by many labs including ours. These include search for: 1) multidomain core enzymes, such as PKS and NRPS, that have unusual domains or domain arrangements; 2) BGCs that contain more

than one core enzymes, which indicate potential convergent assembly of a more elaborate natural product scaffold; 3) BGCs that contain a large number of “tailoring” enzymes, such as oxygenases (P450s, flavin-dependent, nonheme iron-dependent, etc), transferases (acyl, methyl glycosyl), PLP-dependent (transaminases, racemases, β - and γ -replacement enzymes), etc. In addition, clustering of predicted protein products that are hypothetical proteins (HPs) or contain domains of unknown function (DUF) are also strong indicators of potential new enzymology and chemical modifications; and 4) a combination of the above features. In this section, we will summarize a few recent examples of genome mining using these prioritization strategies.

3.1. Unusual core enzyme domain arrangement

3.1.1. HRPKSs that terminate with alternative domains—Structurally diverse polyketides constitute one of the major classes of natural products. In fungi, polyketides are assembled by iterative type I polyketide synthases (PKSs). Each catalytic domain in a multidomain fungal PKS can be repeatedly used to construct the core carbon skeleton. Based on domain architecture and product structures, a fungal PKS can be further classified into either a highly-reducing PKS (HRPKS), a non-reducing PKS (NRPKS), or a partially-reducing PKS (PRPKS).⁹¹ Among them, the programming rules of HRPKSs are the most enigmatic, and the product structures are the most varied. Based on characterized HRPKSs, the release of a completed polyketide chain from the acyl-carrier protein (ACP) can be accomplished by (i) spontaneous α -pyrone formation;^{92,93} (ii) transferring the chain to a partner NRPKS as a starting unit by the starter unit-ACP transacylase (SAT) on the NRPKS;⁹⁴ or (iii) thioesterase (TE)-catalyzed hydrolysis or cyclization.^{95–97} It is rare that a fungal HRPKS employs a NADPH-dependent reductive domain (R) at the C-terminus to reductively release the product, a strategy frequently observed in NRPKS and PKS-NRPS biosynthetic pathways. The first reported example of HRPKS with a R domain mediated product release is Bet1 from the betaenone A (**20**) biosynthetic pathway.⁹⁸ Heterologous expression of Bet1 (HRPKS) and Bet3 (ER) in *A. oryzae* afforded the production of the decalin containing-**21** (Figure 5A). Related compounds to **20** include the sphingolipid synthesis inhibitor australifungin (**22**), of which the terminal β -ketoaldehyde is likely derived from the same reductive release.⁹⁹ In a follow-up study to synthesize a more advanced product derived from **20**, Li et al. reconstituted the biosynthesis of stemphyloxin II (**23**) in *A. nidulans*.¹⁰⁰ A downstream enzyme-catalyzed intramolecular aldol reaction was shown to form the tricycle[6.2.2.0]dodecane structure.

An additional HRPKS-R example was found in calbistrin A (**24**) biosynthesis published by Tao et al.¹⁰¹ The revised structure of **24** consists of a decalin and a polyene that are esterified together. Retrobiosynthetic analysis would have predicted the involvement of two PKSs in the pathway to synthesize the two distinctive polyketide-derived fragments, as seen in the lovastatin biosynthesis.¹⁰² However, comparative genome analysis focused on HRPKSs in three producing fungi led to a candidate *cal* BGC that encodes a single HRPKS.¹⁰³ Heterologous expression of a homologous cluster, *cal'*, was performed in *A. oryzae* to investigate the biosynthesis (Figure 5B). Unexpectedly, the coexpression of CalA' (HRPKS) and CalK' (ER) led to the production of decalin product **25**, while the coexpression of CalA' and CalH' (MT) produced the polyene fragment **26**. This result demonstrates that a single

HRPKS is engaged in producing two distinct products with the aid of different tailoring enzymes. From the structure of **25**, chain release was proposed to be catalyzed by R domain in CalA' via consecutive two-electron reduction to generate an alcohol product. The polyene **26** isolated from CalA'/CalH' coexpression transformant was released as a carboxylic acid, indicating the R domain can only reductively release the decalin product **25**. The use of a single HRPKS in biosynthesis of two different portions of the final polyketide product is reminiscent of the HRPKS in chaetoviridine biosynthesis, in which two triketides of different degrees of reduction are synthesized and intercepted by separate releasing enzymes to build the final product.¹⁰⁴

A single-module NRPS can be fused at the C-terminal of a HRPKS to form a PKS-NRPS megasynthetase. The NRPS module incorporates a nitrogen-containing functional group such as an amino acid, into the natural product. Such fusion of PKS and NRPS can generate a variety of new scaffolds as will be discussed below. A few PKS-NRPS enzymes only contain the condensation (C) module of an NRPS, while the adenylation (A) and thiolation (T) domains are absent. Furthermore, in this PKS-C subgroup, the characteristic HHxxxDG motif of C domain active site is mutated, suggesting noncanonical functions of the C domain. One well-known PKS-C example is lovastatin nonaketide synthase LovB, in which the second histidine in the C domain active site is mutated to an arginine. Based on *in vitro* and yeast reconstitution studies, the proposed function of the C domain is to catalyze the *endo* [4+2] cycloaddition of the hexaketide trienyl intermediate to form the *trans*-decalin.⁵² However, the exact function is still not confirmed, despite an available crystal structure of the standalone C domain,¹⁰⁵ as well as the cryo-EM structure of the entire LovB.¹⁰⁶ Bioinformatics analysis of fungal LovB-like enzymes revealed a clade of PKS-C in which the second histidine residue in the C domain active site is mutated to a proline. To investigate the product of this PKS-C enzyme, Hai et al. expressed one candidate cluster from *Talaromyces wortmanii* in *A. nidulans*.¹⁰⁷ The expression of TwmB (PKS-C), TwmE (ER), and TwmA (TE) resulted in the biosynthesis of the *N*-acylamide compounds wortmanamide A and B (**27** and **28**). Both compounds incorporate the rare ω -amino acid, 5-aminopentanoic acid, that is amidated with a long chain polyketide product (Figure 5C). The function of C domain was further studied by expressing TwmB- C, which is a truncated version of TwmB with only the PKS portion. While no product was detected from TwmB- C alone, coexpression with the standalone C domain restored the production of **27** and **28**. This result implicates the C domain in TwmB is involved in the release of polyketide acyl chain through amide bond formation with ω -amino acid.

3.1.2. Unusual single-module NRPS-like enzymes—NRPSs with noncanonical domain arrangements are also indicative of new functions and new product structures. Canonical NRPS domains include A domains that selectively activate amino acids or other carboxylic acid-containing building blocks, T domains that tether the activated building blocks through thioester bond, and C domains that catalyze amide bond formations. Reductase domains (R) are frequently present at end of NRPS or PKS-NRPS assembly lines to release the product as an aldehyde via NADPH-dependent reduction. Single-module NRPS-like enzymes, such as carboxylic acid reductases (CARs) that contain A-T-R domains, are widely found in bacterial and fungal metabolism. Natural products

containing piperazines, pyridines and morpholines have been associated with such CARs.¹⁰⁸ Given their wide-spread occurrence, CARs are validated leads in genome mining efforts. Schroeder and Keller's group discovered an unusual CAR from *Aspergillus fumigatus*, FsqF, that has an additional PLP-dependent aminotransferase domain appended at the C-terminus.¹⁰⁹ Performing genome mining on the native host identified the BGC, *fsq*, is responsible for biosynthesis of isoquinolines fumisoquins A-C (**29–31**). Isotope feeding studies suggested L-serine and L-tyrosine are incorporated into the fumisoquins. However, [³²P]-ATP-pyrophosphate exchange assay using recombinant FsqF A domain showed neither amino acid is a substrate for FsqF. Experimental data instead suggests that L-serine is first converted to dehydroalanine by the PLP domain, which is then activated by the A domain and thioesterified to the T domain. The carboxylate of L-tyrosine is possibly activated by the ATP-grasp enzyme FsqD, which can be attacked from the dehydroalanyl moiety to form the new C-C bond (Figure 6A). The BGC also encodes a plant-like berberine bridge enzyme, FsqB, that catalyze isoquinoline formation.

Another example of a CAR containing an additional domain is the ATRR enzyme widely conserved in fungi.¹¹⁰ The enzyme group is named ATRR because an additional C-terminal YdfG-like short chain dehydrogenase/reductase (SDR) domain is fused to a CAR, giving the domain architecture A-T-R₁-R₂. This enzyme, initially thought as a candidate to produce a new natural product, is involved in primary metabolism and catalyzes sequential, ATP- and NADPH-dependent reduction of betaine (**32**) to choline (**33**) (Figure 6B). While the oxidation of choline to betaine is known in fungi, the reverse reaction to reduce betaine to choline was not biochemically characterized. Initial inspection of ATRR domain arrangement pointed to a sequential two-electron reductions of a carboxylic acid to an alcohol, with the A domain specifying the identity of the carboxylic acid substrate. Bioinformatics analysis of A domain active site showed mutation of a highly conserved aspartate residue that electrostatically interacts with the α -amino group of amino acids, hence excluding amino acids as substrates. Structural-based prediction using a homology model based on A domain of TycA showed that the A domain in ATRR positions three aromatic amino acids at the end of the active site pocket into an aromatic cage that may interact with a quaternary ammonium group through cation- π interactions. Based on this observation, screening of ATRR activity towards betaine substrates was performed using purified enzyme. Rapid consumption of NADPH was observed in the presence of glycine betaine and ATP. The stepwise reduction was supported by trapping the intermediate glycine betaine aldehyde with phenylhydrazine. In addition, site-directed mutagenesis was performed to generate single R domain mutants. Mutation of R₁ abolished the carboxylic acid reductase activity, while mutation of R₂ abolished the aldehyde reductase activity. Fusion of the second R domain was postulated to enhance substrate channeling in the tandem reduction reaction and to prevent dissociation of the aldehyde that can be readily hydrated. While the genome mining efforts aimed at ATRR did not lead to discovery of a new natural product, the findings provided insights into choline metabolism and betaine homeostasis in fungal organisms.

3.1.3. NRPS-PKS hybrid proteins—Nature has evolved hybrid PKS and NRPKS megasynthetases, such as PKS-NRPS and NRPS-PKS, to generate products with combined

features of polyketide and amino acids. In comparison to PKS-NRPS, NRPS-PKS hybrid enzymes are much rarer in fungal genomes with only a limited number of reports. For example, SwnK, which has domain structure A-T-KS-MAT-KR-ACP-R, was suggested to be involved in swainsonine (**34**) biosynthesis.¹¹¹ SwnK catalysis starts with the activation of pipercolic acid, followed by elongation steps via an intact PKS module. The product is reductively released by R domain which can undergo spontaneous cyclization to complete the indolizidine core (Figure 7A).

A recent example of NRPS-PKS discovered through genome mining was shown to be involved in formation of pyrophen (**35**) and campyrone A (**36**),¹¹² which are amino acid-containing α -pyrone compounds. The BGC encoding an NRPS-PKS was mined from *Aspergillus niger*. In addition to NRPS-PKS (AnATPKS), an *O*-methyltransferase (AnOMT) is colocalized in the cluster, matching the methoxy moiety observed in pyrophen and campyrone A. Heterologous expression of AnATPKS and AnOMT in *A. nidulans* confirmed the cluster is responsible for the formation of **35** and **36**. The NRPS domain activates L-phenylalanine which is used as starter unit of the PKS module. Two decarboxylative Claisen elongation steps followed by a cyclization release step generate the α -pyrones (Figure 7B). Feeding experiment with nonproteinogenic amino acid showed that AnATPKS A domain has relaxed substrate specificity and can accept different aromatic amino acids to arrive at pyrophen analogs.

Tetramate or pyrrolidine-containing natural products from fungi are biosynthesized by PKS-NRPS megasynthetases. After aminoacylation of the β -ketoacyl PKS intermediate, a Dieckmann cyclization catalyzed by an R* domain, or a reductase release by an R domain followed by Knoevenagel condensation, yields either a tetramate or a pyrrolidine, respectively. However, the discovery of tenuazonic acid (**37**) biosynthesis revealed a different strategy to form this core structure from one isoleucine and two acetate units (diketide).¹¹³ BGC for the eukaryotic translation inhibitor **37** was identified from *Alternaria* using comparative RNA sequencing under producing and non-producing conditions. While none of the PKS-NRPS encoding gene displayed change in transcriptional levels, one gene encoding a NRPS-PKS homolog (TAS1) was found to be upregulated under **37** producing conditions. The involvement of TAS1 in **37** biosynthesis was confirmed by knockout. **37** has an *N*-terminal NRPS domain and a single KS domain at the C-terminus. *In vitro* assays treating TAS1 with different CoA substrates revealed that acetoacetyl-CoA is directly incorporated, hence no decarboxylative Claisen condensation is required. Using *N*-acetoacetyl-L-Ile-SNAC to mimic the linear substrate attached to the T domain in NRPS, the authors showed that the standalone KS domain can cyclize the substrate into **37** (Figure 7C). The likely mechanism is KS-catalyzed enolization of the diketone functionality to initiate tetramate formation.

3.2. Multiple core enzyme combinations

To mine compounds with structural complexity, focusing on BGCs containing multiple core enzymes is an attractive option. Molecules derived from such pathways display combined structural features accessible through individual core enzymes. Several modes of core enzyme collaboration have been characterized: 1) one core enzyme is responsible

for biosynthesis of a building block for the other core enzyme. For example, in the biosynthesis of the immunosuppressant drug cyclosporine A (**38**)¹¹⁴, the unnatural amino acid (4*R*)-4-[(*E*)-2-butenyl]-4-methyl-L-threonine (Bmt) incorporated by the cyclosporine NRPS is biosynthesized by a HRPKS in the BGC (Scheme 3A). Release of the polyketide product is subjected to α -oxidation to a ketone followed by transamination to yield Bmt. Another example is in the biosynthesis of ergotamine alkaloids (**39**),¹¹⁵ which involves first the generation of lysergic acid by the actions of numerous enzymes including 4-dimethylallyl tryptophan synthase (4-DMATS), followed by a NRPS that uses lysergic acid as a starter unit to arrive at the final product; 2) sequential biosynthesis of chemically distinct portions of a natural product by different core enzymes. This division of labor is most well-characterized in the collaboration between NRPKS and HRPKSs in the biosynthesis of resorcylic acid lactone (RAL) natural products,¹¹⁶ such as hypomycesin (**40**)¹¹⁷ and zearalenone (**41**).¹¹⁸ In these pathways, the HRPKS generates the highly reduced polyketide chain that is transferred to the NRPKS for elongation and cyclization into the 2,4-dihydroxybenzoic acid moiety; and 3) convergent synthesis of the final product by core enzymes, as exemplified in the biosynthesis of lovastatin (**42**). The lovastatin nonaketide synthase LovB synthesizes the decalin product dihydromonacolin L (DML), while the lovastatin diketide synthase (LovF) synthesizes the α -methylbutyrate side chain that is transferred to the oxidized DML (monacolin J).⁵² In this section, we will highlight some recent examples of genome mining prioritized by the presence of two of more core enzymes in the same BGC.

3.2.1. Multiple PKSs—In addition to RALs, collaborating PKSs in fungi have been shown to synthesize complex polyketides, including sorbicillin (**43**)¹¹⁹ and chaetoviridin A (**44**) (Scheme 3B).¹⁰⁴ In genome mining of the biofertilizer *Trichoderma afroharzianum* t-22, Chen et al. identified an unusual dual PKS BGC (*tlm*) that also encodes a didomain enzyme TlnC. TlnC is a fusion between an *N*-terminal ACP domain and a *C*-methyltransferase (MT) domain.¹²⁰ Sequence analysis of the ACP domain revealed the conserved serine residue for pPant modification is mutated. As a result, the ACP domain is not a functional thioester carrier and was renamed as a pseudo-ACP (ψ ACP). Reconstitution of the BGC in yeast, which involved the coexpression of HRPKS (TlnA), NRPKS (TlnB), TE (TlnD), flavin-dependent monooxygenase (TlnE), *O*-MT (TlnF) and TlnC, led to the biosynthesis of a redox pair of tricholigan A (**45**) and tricholigan B (**46**), both of which are new natural products (Figure 8A). The diene portion of **45** is synthesized as a triketide by HRPKS, whereas the methyl-substituted 2,4-dihydroxybenzoic acid portion is synthesized by the NRPKS in collaboration with TlnC. TlnD is involved in the release of the product from NRPKS. *Ortho*-hydroxylation by TlnE followed by *O*-methylation by TlnB gives the *ortho*-hydroquinone **45**, which can be oxidized to **46**. The role of **45** in the rhizosphere, where *T. afroharzianum* was found, was proposed to reduce extracellular Fe³⁺ to Fe²⁺, which can increase plant uptake during iron-limiting conditions.

The role of the ψ ACP-MT (TlnC) was examined in detail using purified HRPKS, NRPKS and TE. It was determined that the ψ ACP domain of TlnC is involved in recruiting the MT domain to NRPKS for a single C $_{\alpha}$ -methylation step during chain elongation cycle of the NRPKS. Interestingly, when the active site residue in ψ ACP was mutated back

to serine, the ACP domain regained ability to be phosphopantetheinylated, but the MT domain lost the function to perform the programmed methylation. This was attributed to the gain-of-function pPant arm that can enter the *cis*-fused MT active site and block the *in trans* C-methylation of the polyketide intermediate attached to TlnB pPant arm. Taken together, this indicated ψ ACP is important for protein-protein interactions between the megasynthase and *trans*-acting MT domain. This finding was used to mine other BGCs encoding ψ ACP fusion proteins. One such example is a tandem HRPKS and NRPKS-encoding BGC found in *Metarhizium anisopliae*, which encodes a ψ ACP-TE fusion didomain protein.¹²¹ The ψ ACP active site is also mutated from the conserved amino acid sequence DSL to NQI, indicating loss of phosphopantetheinylation. Reconstitution of the two PKSs and the ψ ACP fusion in *A. nidulans* led to biosynthesis of both olivetolic acid (**48**) and the longer sphaerophorolcarboxylic acid (**49**) (Figure 8B). Both compounds are precursors to plant-derived cannabinoids: **48** can be geranylated to cannabigerolic acid (CBGA) and further cyclized into tetrahydrocannabinolic acid in the plant pathways. The *A. nidulans* host, after minimal optimization, was able to produce > 1 g/L of **49**. Mining of additional homologous BGCs led to pathway that can exclusively produce **48**. It is speculated here that ψ ACP recruits the fused TE partner to NRPKS to hydrolytically releases the free carboxylic acid products.

Kaneko et al. found a new HRPKS releasing mechanism through collaboration with a type III PKS.¹²² Genome mining of *Penicillium soppi* revealed a simple but unique BGC, which consists of a HRPKS (PspA), a type III PKS (PspB), and a P450 (PspC). Type III PKSs are ACP-independent KS dimers that can catalyze one or more cycles of decarboxylative Claisen condensation with malonyl-CoA. Fungal type III PKSs are functionally less diverse compared to plant type III PKS.¹²³ Heterologous expression of PspA, or PspA together with PspB in *A. oryzae*, led to the production of two new metabolites, soppline A (**50**) and B (**51**), respectively (Figure 8C). The structural difference between **50** and **51** suggests that PspA synthesizes an unsaturated polyketide molecules, which is transferred to PspB as a starter unit, followed by three cycles of chain extension and cyclization to complete the biosynthesis of alkylresorcinol soppline B. This is the first example of a biosynthetic collaboration between a HRPKS and a type III PKS.

3.2.2. PKS-NRPS with HRPKS—Fungal PKS-NRPSs produce aminoacylated polyketides that can be released as tetramates or pyrrolidone natural products. The A domains in PKS-NRPSs typically activate and incorporate proteinogenic amino acids, such as L-serine in equisetin (**52**),¹²⁴ L-valine in myceliothermophin A (**53**),¹²⁵ L-phenylalanine in leporin B (**54**)¹²⁶ and L-tryptophan in cyclopiazonic acid (**55**)¹²⁷ (Scheme 3C). Many NRPS assembly lines can activate unnatural amino acids, which are biosynthesized by dedicated enzymes encoded in the BGC. Genome mining of an atypical *pox* BGC in *Penicillium oxalicum* encoding both a PKS-NRPS and an HRPKS revealed the first example of collaborative biosynthesis between such a pair of megasynthases.¹²⁸ Transcriptional activation of the *pox* BGC led to the formation of ten related oxaleimide compounds. The most abundant oxaleimide A (**56**) contains a decalin core fused to a succinimide fragment that is substituted with a branched alkyl unit with a terminal olefin (Figure 9). The most bioactive compound is oxaleimide I (**57**), in which the succinimide is oxidized

to a maleimide. Knockout of a hydrolase-encoding gene in the BGC followed by shunt product characterization provided the first clue that the succinimide fragment is derived from the unnatural amino acid (*S,E*)-2-aminodec-4-enoic acid (**58**). Isotopic labeling experiment using [1,2-¹³C₂]-acetate revealed the carbon backbone of **58** is derived from a polyketide pathway. Heterologous expression of the HRPKS PoxF in *A. nidulans* led to the biosynthesis of (*E*)-dec-4-enoic acid. The programming rule of the HRPKS selectively skips one round of enoyl-reduction to retain the olefin. Further coexpression of a P450 enzyme PoxM and an aminotransferase PoxF led to generation of the amino acid **58**, via C_α-oxidation to the ketone and reductive transamination, respectively (Figure 9B). The amino acid **58** is then activated by the A domain in the PKS-NRPS PoxE and aminoacylated with the acyclic polyketide acyl intermediate. Thus, the role of HRPKS is to synthesize a nonproteinogenic amino acid for the NRPS module in the PKS-NRPS, which draws parallel to formation and incorporation of Bmt into cyclosporine.¹¹⁴ After reductase release of the acyclic aminoacylated polyketide product as an aldehyde, [4+2] cycloaddition by the Diels–Alderase forms the decalin core and Knoevenagel condensation by the hydrolase forms the pyrrolidone. Oxidation of the pyrrolidone then triggers a rearrangement that leads to migration of the allyl group observed in the final product **57**. The precise position of the olefin in the amino acid **58** is key to this semipinacol-like rearrangement that forms the succinimide structure in this family of compounds.

3.2.3. Terpene Cyclase with NRPS—Terpenoids represent the largest class of natural products. The carbon backbones of terpenoids are polymerized and cyclized from isopentenyl building blocks (IPP, DMAPP) by terpene cyclases (TCs).¹²⁹ The cyclized hydrocarbons can then be subjected to a multitude of cation-mediated rearrangements and oxidations to give diverse bioactive products. Terpene-polyketide hybrid natural products, often named meroterpenoids, are isolated from both bacteria and fungi. The biosynthesis and genome mining of these compounds have been well-documented.¹³⁰ In contrast, BGCs encoding both TCs and NRPSs have not been studied extensively. Here we are excluding BGCs in which prenyltransferases (PTs) are coexpressed to transfer geranyl, farnesyl or geranylgeranyl units to NRPS-derived core structures, but rather are referring to those that express TCs and NRPSs that can potentially lead to terpenoid-amino acid or terpenoid-alkaloid hybrid compounds. One of the first example of such a BGCs was reported by Lee et al., who elucidated the biosynthetic pathway for aculene A (**59**), a norsesquiterpenoid (14-carbon) that is aminoacylated with L-proline.¹³¹ Scanning the producing host led to identification of the *ane* BGC that encodes a single module NRPS, a TC and a set of oxidative tailoring enzymes. Heterologous expression of the BGC in *A. oryzae* confirmed the cluster is linked to aculene A and related compounds. Using a bottom-up approach, a P450 enzyme in the BGC was found to be responsible for decarboxylation to yield the norsesquiterpene core, while the NRPS is involved in L-proline activation and aminoacylation of the oxidized terpenoid (Figure 10A).

To explore the biosynthetic diversity of TC-NRPS hybrid pathways, Yee and Kakule performed genome mining and showed such hybrid clusters are in fact wide-spread in fungi. Focusing on the *flv* cluster from *Aspergillus flavus*, a heterologous expression approach using *A. nidulans* was performed. The *flv* cluster is particularly interesting

in that it contains two predicted TCs (FlvE and FlvF) and one single-module NRPS (FlvI).¹³² In addition, numerous accessory enzymes are encoded in the BGC, including two P450s, an SDR (FlvB), an ornithine decarboxylase (FlvG), and a didomain enzyme (FlvA) with a non-heme iron-dependent (NHI) oxygenase fused to a PLP-dependent enzyme. Heterologous expression of the entire BGC led to biosynthesis of a new natural product flavunoidine A (**60**), a tripartite molecule (Figure 10B). The heavily oxidized terpenoid portion is synthesized by FlvE and oxidized by one of the P450 (FlvD). Surprisingly, a dimethylcadaverine substituent is connected to the terpene core via an axial C-N bond. Systematic reconstitution efforts showed that dimethylcadaverine is synthesized from L-lysine by FlvG and a cryptic methyltransferase (FlvH) initially annotated as an HP. The second TC in the BGC, FlvF, is responsible for formation of the C-N bond, only in the presence of the P450 enzyme FlvD. In the absence of FlvF, the terpenoid core is instead connected via C-N bond to ethanolamine in both axial and equatorial configurations, suggesting nonenzymatic quenching of a possible cation intermediate. The proposed mechanism is the P450 FlvD oxidizes the terpenoid core in sequential one-electron steps to generate a secondary carbocation, which is intercepted by FlvF with dimethylcadaverine. Therefore, FlvF does not function as a *bona fide* TC, but may be responsible for stereoselective C-N bond formation that is not seen in terpenoid maturation.¹³³ The third part of flavunoidine is a new-to-nature 5,5-dimethyl-L-pipecolate (**61**) esterified to a hydroxyl group introduced by the P450 FlvE in the terpenoid core. This unusual amino acid is synthesized from the tandem actions of FlvA and FlvB, of which a key step is proposed to be the γ -replacement reaction using *O*-acetyl-L-homoserine as the latent vinyl glycine donor (see section 3.3.2). Finally, the NRPS FlvI adenylates and esterifies **61** to the terpenoid to generate flavunoidine. The unexpected structural features of flavunoidine showcases how TC-NRPS hybrid pathways can afford new and complex natural products.

The pyrrolobenzoxazine terpenoid CJ-12662 (**62**) from *Aspergillus fischeri* var. *thermomutatus* ATCC 18618 is another example of such fusion between terpene and amino acid. A known compound with potent anthelmintic activity, CJ-12662 contains two distinct substructures that are esterified together, indicating NRPS involvement. The terpenoid portion is a heavily hydroxylated amorpho-4,11-diene, while the pyrrolobenzoxazine portion is clearly derived from oxidative rearrangement of L-tryptophan. Genome scanning identified the candidate *thm* BGC, which was reconstituted using *A. nidulans* to confirm its role in CJ-12662 biosynthesis (Figure 10C).¹³⁴ Amorpho-4,11-diene, a famous plant terpene that is the precursor to artemisinin, is synthesized by the TC ThmB and triply hydroxylated by three P450 enzymes in the BGC. The resulting triol is esterified regioselectively at the C₂ alcohol with L-tryptophan by the NRPS ThmA. The indole portion is then oxidatively modified to pyrroloindole via the action of a flavin-dependent epoxidase, followed by *N*-methylation, chlorination, and *N*-oxidation to trigger a [1,2]-Meisenheimer rearrangement and give CJ-12662. A related pathway involved in biosynthesis of an epoxidized version of CJ-12662 was independently reconstituted in *A. nidulans* by Hu and coworkers.¹³⁵ The authors in this study suggested a possible role of the *N*-oxygenase in facilitating the Meisenheimer rearrangement.

3.2.4. Cryptic Terpene Cyclase with NRPKS—Xenovulene A (**63**), a potent inhibitor of the GABA-benzodiazepine receptor, is an unusual meroterpenoid isolated from *Acremonium strictum* IMI 501407 (currently verified as *Sarocladium schorii*).¹³⁶ The structure of **63** contains a rearranged polyketide-derived moiety fused with a humulene sesquiterpene. It was hypothesized that the cyclopentenone is formed from methylorsellinic acid via a tropolone intermediate. Formation of tropolone in fungi is known through studies on stipitatic acid (**64**): biosynthesis of orsellinic acid or aldehyde by a NRPKS (TropA) is followed by tandem oxidative modification catalyzed by FAD-dependent enzyme (TropB) and nonheme iron-dependent (NHI) dioxygenase (TropC).¹³⁷ Sequencing the producing strain of **63** led to a candidate BGC (*aspks1*) that encoded these three genes required for tropolone biosynthesis, as well as other tailoring enzymes. Heterologous expression in *A. oryzae* confirmed the minimal set of genes responsible for the biosynthesis of **63**, including NRPKS (*aspks1*), an FAD-dependent hydroxylase (*asL1*), a NHI dioxygenase (*asL3*), a P450 (*asR2*), two proteins with unknown functions (*asR5* and *asR6*), and two putative NAD/FAD-dependent oxidoreductases (*asL4* and *asL6*) (Figure 11A).¹³⁸ Although α -humulene is reported as a fungal metabolite synthesized by a class I TC,¹³⁹ no homologous protein can be found in *S. schorii*. After coexpression of different combinations of genes, the authors confirmed that *asR5* and *asR6* are involved in the production of α -humulene and the proposed cycloaddition between α -humulene and polyketide product, respectively.

To investigate the function of AsR5 and AsR6, recombinant proteins were obtained from *E. coli*. Incubation of AsR6 with farnesyl pyrophosphate (FPP) and Mg^{2+} led to the production of α -humulene, indicating the function of AsR6 is indeed a TC. Although sequence alignment with type I terpene cyclase indicated AsR6 lacks well-conserved magnesium binding residues, the enzyme is Mg^{2+} -dependent as confirmed through biochemical assays.¹³⁸ The fusion between tropolone and α -humulene was proposed to occur through an intermolecular hetero-Diels–Alder reaction. In a follow-up study, Hu and coworkers functionally characterized the homolog of AsR6, EupfF, in the BGC of related compound eupenifeldin (**67**).¹⁴⁰ Dehydration of the tropolone to form a reactive α -quinone methide (**65**) followed by hetero-Diels–Alder reaction with hydroxyl-humulene led to the product neosetophomone B (**66**) that can be further processed into eupenifeldin (Figure 11B). Therefore, the BGCs of **63** and **67** contain two cryptic enzymes: a sesquiterpene cyclase and a pericyclase, that were functionally reconstituted in heterologous host. Cox and coworkers subsequently reported the combinatorial biosynthesis using *A. oryzae* as heterologous host.¹⁴¹ Coexpression of genes from three *arpks1* homologous clusters led to the generation of unnatural tropolone sesquiterpenoids. Recently, the total synthesis of the related pycnodione and DFT calculations led to the conclusion that the second hetero Diels–Alder reaction observed in such bistropolone sesquiterpenes must also be enzyme-catalyzed.¹⁴² The responsible enzyme and substrates have not been identified to date.

3.3. Combinations of accessory enzymes

One strategy used by the community to mine known-unknown BGCs is focusing on those that encode a multitude of accessory enzymes in addition to a core enzyme. Clustering of oxidative enzymes such as P450s, NHI oxygenases, and/or flavin-containing monooxygenases (FMOs), etc. is usually correlated with extensive structural modifications.

Examples of such modifications have been discussed in some of the previous examples, and additional examples will be presented in 3.3.1. Other accessory enzymes such as transferases, hydrolases are also scaffold modifying, and often function in series with the oxidative enzymes. In recent years two additional classes of enzymes have emerged to be strong indicators of chemical complexity. The PLP-dependent enzymes will be discussed in 3.3.2, while the pericyclases family of enzymes will be discussed in section 4.

3.3.1. Multitude of redox enzymes—Comparison of homologous BGCs can reveal variations in the number of biosynthetic enzymes, including oxidative enzymes. This in turn can lead to mining of natural product variants with potentially more complex structures. One example is the discovery of the ilicicolin H analog Py-469 (**68**) from *Penicillium variable* through the combination of genome mining and microcrystal electron diffraction (MicroED).¹⁴³ Ilicicolin H (**2**) is a 2-pyridone fungal natural product that is a potent inhibitor of eukaryotic respiratory chain.^{144,145} The biosynthesis of ilicicolin requires a PKS-NRPS, an enoylreductase (ER) and ring expansion P450 (P450_{RE}) that together generate the tyrosine-derived 2-pyridone.³³ The poly-olefinic portion of the compound is cyclized by a pericyclase IccD in an inverse-electron demand Diels–Alder reaction to form the decalin.³³ The final step is an epimerization catalyzed by a flavin-dependent enzyme to arrive at ilicicolin H. This pathway was fully reconstituted in *A. nidulans* by coexpressing the five enzymes involved. Comparative analysis showed these five genes are well-conserved in several fungal species. Interestingly, the *icc* cluster in *P. variable* encodes three additional genes, which are a P450 (IccF), a SDR (IccH), and a flavin-dependent oxidoreductase (IccG). It was hypothesized that these three enzymes can catalyze further modifications on the ilicicolin H structure. Upon coexpressing the three genes together with the five that produced ilicicolin H, a new metabolite **68** was observed. NMR experiments showed the hydroxyphenyl ring derived from tyrosine has been modified into the 2,3-epoxy-1,4-cyclohexanediol by the three enzymes, via a cascade of hydroxylation, epoxidation and ketoreduction (Figure 12A). The relative stereochemistries of the epoxy-cyclohexanediol, as well as the stereochemical relationship to the decalin ring system, however, cannot be determined by NMR analysis in part due to the rigid, flat 2-pyridone ring system. MicroED, however, enabled rapid determination of the relative stereochemistry of **68**. This example shows the combination of heterologous expression and MicroED structural determination can greatly accelerate natural product genome mining efforts.

Another genome mining example involving a BGC with a multitude of redox enzymes is that of trichoxide from *Trichoderma virens*.¹⁴⁶ Tabulation of the HRPKS-encoding BGCs from this well-studied fungal host revealed the *vir* cluster that is particularly interesting. In addition to the HRPKS VirA, the BGC encodes potentially nine redox enzymes, including four SDRs, one P450, two flavin-dependent oxidoreductase and two cupid-domain containing oxidoreductases. Heterologous expression of the entire cluster in *A. nidulans* led to the discovery of the new metabolic trichoxide (**69**), which is a new member of the epoxy-cyclohexenol containing natural products. Bottom-up reconstitution starting with the HRPKS showed the biosynthetic pathway can be divided into two parts, the formation of an aromatic salicylaldehyde, and dearomatization to the epoxy-cyclohexanol (Figure 12B). Formation of an aromatic intermediate in a HRPKS-containing BGC was surprising,

considering HRPKSs generate highly reduced compounds. Analysis of shunt products and intermediate, however, revealed the biosynthetic strategy. VirA synthesizes a reduced polyketide in which the last three ketides are reduced to β -hydroxyl groups. The triol is released reductively to an aldehyde, which can cyclize into a cyclic hemiacetal. Stepwise oxidations of two of three β -alcohols yield a poly- β -ketone intermediate (**70**) that more resembles a NRPKS product. Intramolecular aldol cyclization followed by dehydration and aromatization give the salicylaldehyde intermediate (**71**). This unexpected logic of using an HRPKS in combination with redox enzyme to generate salicylaldehyde was also seen in the biosynthesis of the mycotoxin sordarial (**72**) from *Neurospora crassa*, using *A. nidulans* as a heterologous host.¹⁴⁷ From the salicylaldehyde, the remaining redox enzymes dearomatize the aromatic ring through a series of hydroxylation, epoxidation and reduction steps, a sequence of reactions similar to that take place during biosynthesis of **68** (Figure 12A).

One additional notable example of how redox enzymes can heavily modify a core scaffold is in the biosynthesis of novofumigatonin (**73**), which is isolated and characterized from *Aspergillus novofumigatus* IBT 16806.¹⁴⁸ Novofumigatonin is a dramatically oxygenated meroterpenoid containing an orthoester group, two lactone rings, and an aldehyde group. Orthoester is a unique functional group that contains three alkoxy groups attached to a single carbon atom. The classic example of an orthoester natural product is tetrodotoxin (**74**) (Figure 12C).¹⁴⁹ The precursor of novofumigatonin was proposed to be derived from asnovolin A (**75**), which is synthesized from the aromatic polyketide 3,5-dimethylorsellinic acid and the C15 isoprenoid farnesyl-diphosphate (Figure 12D). To solve the biosynthetic pathway of **73**, Matsuda et al. employed CRISPR-Cas9 based gene deletion, heterologous expression in *A. oryzae*, and in vitro analysis. The oxidative decoration begins from the NvfK-catalyzed epoxidation of farnesyl-dimethylorsellinate to initiate cyclization of the tetracyclic asnovolin H (**76**). The C₃-OH is then oxidized to a ketone by NvfC and expanded to the seven-member lactone by a Baeyer-Villiger type of oxygenase NvfH. Formation of the hydroxy endoperoxide in fumigatonoid A (**77**) is catalyzed by a nonheme, Fe(II)/ α -ketoglutarate-dependent enzyme, NvfI, consuming two molecules of oxygen: the first oxygen is incorporated intact as the bridging endoperoxide, while an oxygen atom from the second molecular oxygen is incorporated as a hydroxyl group at C₃' position via a radical rebound step. Formation of the lactone and reduction of C₄' ketone affords fumigatonoid B (**78**). Two additional NHI enzymes, NvfE and NvfF, complete the biosynthesis of novofumigatonin with the orthoester moiety. NvfE is responsible for the formation of orthoester moiety in fumigatonoid C (**79**) with no change in oxidation state, indicating that it functions as an isomerase instead of an oxygenase. Further in vitro analysis showed that NvfE is a cofactor- and cosubstrate-free enzyme with mutation of the highly conserved glutamate required for α -ketoglutarate binding. The last enzyme NvfF is a *bona fide* Fe(II)/ α -KG-dependent dioxygenase that catalyzes orthoester exchange and oxidation of the released alcohol to aldehyde. Overall, three out of eight oxygen atoms in novofumigatonin are from the orsellinic acid building block. The other five oxygen atoms are incorporated from four molecular oxygens by a series oxygen-handling enzymes, showcasing how nature generates a remarkable set of oxygen functional groups in a compact framework.

3.3.2. PLP-dependent enzymes—PLP-dependent enzymes are synonymous with amino acid metabolism. The use of PLP enzymes in natural product biosynthesis is also well-documented.¹³⁰ These enzymes are associated with substrate decarboxylation, racemization and formation of C-C bonds, the latter catalyzed by PLP-dependent threonine aldolases.¹⁵⁰ One family of PLP-dependent enzyme with sequence homology to cystathionine γ -synthase has recently emerged in both bacterial and fungal BGCs to catalyze interesting C-C, C-O and C-N bond forming steps through a γ -replacements reaction involving *O*-acylated-L-homoserine. Van Lanen and coworkers noted the first example of a C-N bond forming, γ -replacement reaction catalyzed by a PLP-enzyme in nucleoside antibiotic biosynthesis in bacteria.¹⁵¹ Around the same time, three different C-C bond formation enzymes in biosynthesis of substituted pipercolate amino acids were found in fungi through genome mining efforts, including the FlvA example involved in 5,5-dimethyl-pipercolate formation during flavunoidine biosynthesis (Figure 10B). Here the other two examples will be briefly discussed.

The first example of a PLP-dependent enzyme catalyzing γ -replacement, C-C bond forming reaction is CndF found in the citrinadin BGC from *Penicillium citrinum*.¹⁵² Citrinadin A (**80**) and B (**81**) (Figure 13A) are prenylated indole alkaloids derived from a tryptophanyl-6-methyl-pipercolate ketopiperazine biosynthesized by a bimodular NRPS. The nonproteinogenic (2*S*, 6*S*)-6-methyl-pipercolate (**82**) is not a previously known metabolite and must therefore be biosynthesized from dedicated enzymes encoded in the *cnd* BGC. Comparative analysis with other BGCs of indole alkaloids that do not incorporate **82** in the scaffold suggested three *cnd* enzymes, CndE (SDR), CndF (PLP-dependent enzyme with 36% sequence identity to cystathionine- γ -synthase), and CndG (HMG-CoA lyase), may be involved. Heterologous expression of the three enzymes in *A. nidulans* indeed led to the accumulation of **82** (Figure 13B). Exclusion of CndG in the heterologous host did not abolish the biosynthesis of **82**, albeit led to decreased levels. Since CndG is predicted as an HMG-CoA lyase that can cleave HMG-CoA into acetyl-CoA and acetoacetate, it was hypothesized that acetoacetate could be the three-carbon nucleophile used by CndF in the PLP-dependent γ -replacement reaction. Feeding experiment using [2,4-¹³C₂] ethyl acetoacetate, which can be hydrolyzed by endogenous esterase in *A. nidulans* expressing CndE and CndF, resulted in C₅ and C₆-methyl doubly labeled **82**. Based on additional biochemical evidences, CndF is proposed to catalyze C-C bond formation between acetoacetate and enzyme-bound ketimine form of vinyl glycine to give (*S*)-2-amino-6-oxoheptanoate (**83**), which upon release from the enzyme can cyclize to form the Schiff base (**84**). The SDR CndE then catalyzes stereospecific imine reduction to afford **82**. Encouraged by the finding that β -keto carboxylates such as acetoacetate can be generated *in cellulo* by endogenous esterase from ethyl esters, *A. nidulans* expressing CndE and CndF was used as a biotransformation host to convert β -keto ethyl esters to 6-alkyl pipercolate derivatives. These studies showed CndF has considerable promiscuity towards α - and γ -substituted β -keto carboxylate compounds, including bulky and cyclic substrates. Thus, CndF may be further repurposed as a biocatalyst to construct 6-alkyl pipercolate derivatives.

In a parallel study, a two-enzyme combination, including a PLP-dependent enzyme (Fub7) and an FMN-dependent oxidase (Fub9), was discovered in fusaric acid (**85**) biosynthetic pathway to construct 5-alkyl substituted picolinic acid.¹⁵³ Fusaric acid is an inhibitor of dopamine β -hydroxylase produced by *Fusarium* species, and BGC was previously identified through genetic knockout experiments.¹⁵⁴ A 6π electrocyclization step was proposed to generate the picolinic acid moiety. Interestingly, the BGC contains a HRPKS (Fub1) and a NRPS-like CAR (Fub8), as well as a few other enzymes including Fub7 and Fub9. Heterologous expression in *A. nidulans* was performed to clarify the biosynthetic route as shown in Figure 14A. Coexpression Fub1, Fub4 (α/β hydrolase), Fub6 (reductase), Fub7, Fub8 and Fub9 led to biosynthesis of **85**, establishing the minimal set of enzymes needed. *In vitro* reconstitution showed that Fub1, 4, 6 and 8 collectively synthesizes *n*-hexanal, a rather simple aldehyde substrate that requires the collaborative functions of an HRPKS (carbon backbone generation) and a CAR (reductive release as aldehyde). The α -carbon of *n*-hexanal turned out to be the carbon nucleophile in the Fub7-catalyzed γ -replacement C-C bond formation using *O*-acetyl-L-homoserine as the latent electrophile. The resulting (2*S*)-2-amino-5-formylnonanoic acid (**86**), upon release from the PLP enzyme, can cyclized into the Schiff base tetrahydrofusaric acid (**87**). Fub9 was confirmed to catalyze the four-electron oxidation to generate **85**. Similar to CndF, Fub7 showed substantial promiscuity towards aldehyde substrates with different carbon length and substituents. Combining Fub7 and Fub9 activities, a panel of 5-alkyl, 5,5-dialkyl, and 5,5,6-trialkyl-pipcolic acids were prepared. PLP-enzymes with sequence homology to CndF and Fub9 are widely found in fungal BGC and can serve as new leads to in genome mining efforts.

A bifunctional PLP-dependent enzyme was discovered to work together with an HRPKS, leading to the discovery of new biosynthetic logic to form the indolizidine scaffold.¹⁵⁵ The indolizidine structure is commonly found in various plant alkaloids, including vinblastine and vincristine. An early study of indolizidine formation in fungi was that of swainsonine (**34**) biosynthesis.¹¹¹ It was reported that the formation of indolizidine requires the reductive release of polyketide chain from a NRPS-PKS assembly line (Figure 7A). However, the same biosynthetic logic cannot rationalize the biosynthesis of curvulamine (**88**) (Figure 14B). Dai et al., discovered a new indolizidine forming pathway through a collaboration between a HRPKS and a PLP-dependent enzyme. The putative BGC was identified from analyzing RNA expression levels and verified with gene inactivation. The BGC encodes a HRPKS (CuaA), a PLP-dependent aminotransferase (CuaB), a SDR (CuaC) and a FMO (CuaD). Heterologous expression of CuaA and CuaB in both *A. oryzae* and yeast led to accumulation of **89**, which contains a 3H-pyrrol-3-one structure with hydroxylation at C₉ position. Further coexpression of CuaD led to epoxidation of **89** to **90**. Biochemical characterization showed that CuaB alone can catalyze formation of **89** when using a SNAC-tethered polyketide substrate mimic. The origin of the C₉ hydroxyl group was confirmed to derive from molecular oxygen, leading to a proposed mechanism in which a resonance stabilized carbanion activates molecular oxygen to perform substrate oxidation, which has been noted for PLP-dependent enzymes in natural product biosynthesis.¹⁵⁶ CuaB is therefore a bifunctional PLP-dependent enzyme in biosynthesis of curvulamine. It catalyzes a Claisen condensation to release the polyketide chain from a HRPKS, followed by substrate-hydroxylation in a paracatalytic reaction.

3.4. Other considerations

The above examples rely on the clustered nature of biosynthetic genes, in which a BGC encodes all the necessary proteins and enzymes for biosynthesis of target compound. In fungi, this appears to be overwhelmingly the case, which has greatly facilitated biosynthetic reconstitution and genome mining. However, given the intrinsic difficulty in finding unclustered genes, the occurrence of unclustered BGCs is surely underestimated. One example is that of the meroterpenoid austinol (**91**) discovered by the Wang group.¹⁵⁷ The BGC of **91** was initially searched for by using PKS and prenyltransferase as targets. However, no such enzyme combination was found to be colocalized in *A. nidulans* genome. An intensive gene inactivation experiment that afforded 20 different mutant strains was carried out to identify gene candidates involved in biosynthesis of **91**, which were located in two different BGCs on the *A. nidulans* genome.

In the case where a single enzyme is unclustered with the BGC, heterologous biosynthesis can serve as a facile approach to screen potential candidates. Such a case study can be found in the identification of unclustered terpene cyclases in the biosynthesis of fungal indole diterpenes, such as aflavinine (**92**) and anominine (**93**) (Scheme 4).⁵⁵ These compounds are antiinsectants and are produced in the sclerotia of several *Aspergillus* species. All biosynthetic genes for related compounds such as paxilline (**5**) and aflatram (**94**) are clustered, which include a geranylgeranyl diphosphate synthase (GGPPS), a GGPP transferase that forms geranylgeranyl indole, multiple epoxidases and an indole diterpene cyclase (IDTC). However, no gene encoding an IDTC was clustered with the other biosynthetic genes in the hosts that produce **92**, **93** and tubingensin A (**95**). A phylogenetic analysis of all the IDTC homologues found in the genomes of the producing strains led to the identification of numerous candidates. Using yeast as a heterologous host, the IDTC candidates were individually coexpressed with the other biosynthetic enzymes followed by metabolite analysis. This led to the verification of AfB and AtS2B as unclustered IDTCs in the pathways of anominine and aflavinine, respectively. The evolutionary reason for such unclustered feature is unknown. One possible explanation is that these IDTCs are co-regulated with sclerotia-specific genes and are not co-regulated with the upstream enzymes in the BGCs that synthesize the uncyclized indole diterpene. Regardless of the biological intentions, one should keep in mind the possibility of unclustered biosynthetic genes in genome mining efforts.

On the other hand, duplication of biosynthetic genes into multiple BGCs in the same host has also been noted and should be considered in genome mining. One such example is in the biosynthesis of the heptacyclic duclauxin (**96**) from *Penicillium duclauxi* (Figure 15).¹⁵⁸ Duclauxin belongs to fungal aromatic polyketides that has a characteristic 6/6/6/5/6/6/6 ring system derived from the dimerization of oxaphenalenones. Structural variations among duclauxin derivatives arise from different oxaphenalenone monomers and dimerization regiospecificities. The precursor of oxaphenalenone monomers, phenalenone (**97**) is synthesized by an NRPKS and oxidatively rearranged by a FMO.¹⁵⁹ In the producing strain *Talaromyces stipitatus* ATCC 10500, the *dux1* cluster was identified using NRPKS and FMO as BLAST queries. Gene deletion of NRPKS (*dux1*), completely abolished the production of **96**. Remaining redox enzymes encoded in the *dux1* cluster include two P450s

(DuxD and DuxL), oxidoreductase (DuxB), NAD(P)H-dependent reductase (DuxA, DuxG, and DuxJ), and a cupin family oxygenase (DuxM), which were proposed to be essential in transforming phenalenone to oxaphenalenone. While inactivation of some of these, such as *duxH* and *duxM* led to the complete abolishment of **96**, other knockouts such as *duxA* and *duxL* had no effects. Detailed BLAST analysis revealed a separate cluster, *dux2*, encoding DuxA', DuxB', DuxC', DuxD', DuxG', DuxH', DuxL', and DuxM', all of which showed over 70% sequence identity to the corresponding proteins in *dux1* cluster. RT-PCR analysis further confirmed that both clusters were expressed when the production of **96** was detected in *T. stipitatus*. Because of such gene duplication of the *dux* clusters, heterologous expression turned out to be the more informative approach to understand individual *dux* gene functions. Stepwise reconstitution in both yeast and *A. nidulans* solved the pathway of **96**, highlighting DuxM is the key enzyme for diversifying the structures of oxaphenalenone monomers. DuxM initiates the redox reaction sequence by an oxidative cleavage of phenalenone through a transient hemiketal-oxaphenalenone intermediate (**98**). DuxM can further catalyze decarboxylation of **98** to give the anhydride **100**, while three other enzymes (DuxJ, DuxD, and DuxG) in the biosynthetic pathway can morph **98** into **99**. Lastly, P450 DuxL catalyzes the oxidative heterocoupling of **99** and **100** to form cryptoclauxin (**101**), or homocoupling of **99** to duclauxin (**96**) (Figure 15).

4. Shunt product formation during heterologous reconstitution.

Shunt product or intermediate? That is the question asked by nearly all researchers working on biosynthesis of natural products. Regardless of using native hosts or heterologous hosts, a change in metabolite profile following either a gene knockout or gene introduction is an exciting result. Identifying an on-pathway intermediate is a desired outcome in reconstruction of the biosynthetic pathways. However, off-pathway shunt products are frequently formed due to chemical instability of intermediates or crosstalk between targeted pathway and endogenous metabolism. In most cases, formation of shunt products presents an obstacle in pathway reconstitution because downstream enzymes cannot recognize the compound to advance the pathway. Incorrect assignment of a shunt product as a biosynthetic intermediate can significantly derail the investigation and lead to misinterpretations of the chemical logic. On the other hand, careful analysis of shunt product structures can provide revealing chemical insights into biosynthetic pathways. One such example is from studying the iterative HRPKSs from fungi, typified by LovB from the lovastatin pathway. Shunt product characterization from the bottom-up reconstitution in heterologous hosts, as well as from interruption of the HRPKS domain functions, revealed highly complex programming rules in chain length control, permutative reductive tailoring and methylations.^{52,160} HRPKS programming will not be elaborated further here, and the readers can refer to the excellent review by Cox that is also in this volume.¹⁶¹ In this section, we will discuss recent representative case studies of how identification of shunt products has impacted the biosynthetic mining and investigation of fungal natural products.

4.1. Shunt products formed from cellular crosstalk

Shunt products can form from crosstalk of target pathway with endogenous enzymes, especially those involved in detoxification of reactive compounds such as aldehydes and

α,β -unsaturated Michael-acceptors, etc. Members of the alcohol dehydrogenases (ADHs), aldehyde dehydrogenases and short-chain dehydrogenase/reductases (SDRs) families from fungi have promiscuous substrate specificities and can rapidly reduce aldehydes into the corresponding primary alcohol. In fungal biosynthetic pathways, one route to generate an aldehyde intermediate is through the reductive release of thioesters by R domains appended at the end of assembly lines (see section 3). In PKS-NRPS assembly lines terminated with R domains, the released aldehyde is immediately cyclized via Knoevenagel condensation with the α -carbon of the 1,3-diketo portion of the polyketide chain to yield a pyrrolidine ring. Although this reaction can occur spontaneously in some pathways, a dedicated hydrolase-like enzyme is required in most studied pathways.^{128,162} In the absence of the hydrolase, the aldehyde is readily reduced to the shunt product alcohol. This was observed in cytochalasin reconstitution in *A. oryzae* and in mining of oxaleimides in *A. nidulans*.^{69,128,163} In the oxaleimide example, which was discussed in detail above, structure of the alcohol shunt product led to clarification of the biosynthetic origin of the unusual succinimide group. One strategy found in fungal biosynthesis to counter the propensity of aldehydes to be reduced to alcohol shunt products is the presence a flavin-dependent oxidase that oxidizes the alcohol to the aldehyde. This strategy was uncovered in the biosynthesis of pyriculol (**102**) which is a devastating mycotoxin. Genome mining of a HRPKS-containing pathway in *Neurospora crassa* led to the biosynthesis of sordarial (**72**), a related salicylaldehyde.¹⁴⁷ In analogous steps to the trichoxide example in Figure 11B, the salicylaldehyde is synthesized by the HRPKS and a cadre of redox enzymes. However, the aldehyde (**103**) is readily reduced to yield the salicylic alcohol (**104**) in *A. nidulans* (Figure 16A). Introduction of a flavin-dependent oxidase, however, restored the formation of the salicylaldehyde that is further processed into sordarial. This is also the role of a homologous oxidase in the pyriculol pathway in *Magnaporthe oryzae*, as knockout of this enzyme led only to biosynthesis of the nonvirulent dihydroxypyriculol (**105**).

The reduction of α,β -unsaturated moieties by endogenous ene-reductases is another common occurrence that can hamper heterologous reconstitution efforts. Ene-reductases are flavin-dependent enzymes that can perform a 1,4-addition with a hydride to yield a saturated shunt product. This was most clearly demonstrated in the attempted reconstitution of UCS1025A (**106**) in *A. nidulans*.¹⁶⁴ UCS1025A is a pyrrolizidinone (azabicyclo [3.3.0] octanone)-containing natural product and is a strong telomerase inhibitor.¹⁶⁵ The pyrrolizidinone is fused with a γ -lactone to give a furopyrrolizidine that is further connected to a polyketide derived *trans*-decalin. While UCS1025A was initially isolated from *Acremonium* sp. KY4917 and a putative gene cluster was located, difficulties involved in growing *Acremonium* sp. and subsequent molecular biology experiments, led to the mining of a homologous cluster found in *Myceliophthora thermophila*. Transcriptional factor overexpression led to production of UCS1025A in this host and confirmed the metabolite of the *ucs* cluster. The pyrrolizidinone ring is generated by the function of a PKS-NRPS, which aminoacylates the polyketide with (4*S*, 5*S*)-4-methylproline. Reductive release and Knoevenagel condensation afford the pyrrolizidinone ring system. However, attempts to characterize the oxidative transformations to furopyrrolizidine using heterologous hosts such as yeast was not successful. This was primarily due to the facile reduction of the ene moiety in pyrrolizidinone to the saturated product. Although these shunt products cannot

be elaborated into UCS1025A, the 6*S*-methyl group in an early shunt product **107** can be oxidized by a P450 enzyme UcsK into the corresponding carboxylic acid (**108**), which confirmed the role of UcsK in the pathway. Characterization of the shunt products also led to an improved biosynthetic pathway proposal in which an oxa-Michael cyclization involving the carboxylate group forms the final product (Figure 16B).

Cellular crosstalk can also come from nonenzymatic reactions between the reactive warhead of the target natural product and endogenous metabolites, via a possible mechanism of cellular detoxification. One such example is in the heterologous biosynthesis of restricticin (**109**).¹⁶⁶ **109** is a potent antifungal natural product that inhibits lanosterol 14 α -demethylase (CYP51), which is a validated antifungal target. A putative gene cluster for restricticin was found in *Aspergillus nominus*, a strain not known to produce the compound, using the SRE-guided approach. The *rstn* cluster encodes a HPRKS that is responsible for synthesizing carbon skeleton of the polyene-tetrahydropyran unit, a single-module NRPS that esterifies the hydroxylated tetrahydropyran with glycine, a set of accessory enzymes and a self-resistance version of CYP51. Heterologous expression of the cluster in *A. nidulans* in CDST media led to formation of trace amounts of restricticin, with the major product being the *N*-acetylated restricticin. *N*-acetylated **109** is significantly weaker as an antifungal since the primary amine in **109** is involved in coordinating to the heme iron in CYP51. The amine, however, is modified in a more surprising fashion when the heterologous strain was grown in the minimal CD media in which no restricticin was observed in the extract. Instead, an adduct (**110**) between restricticin and the *A. nidulans* metabolite aspernidine (**111**) was formed, in which the free amine is incorporated as part of the isoindolone ring. The adduct is proposed to form from the attack of restricticin on the quinone methide precursor (**112**) of **111**, followed by cyclization (Figure 16C). In comparison, **111** is formed when **112** is attacked by ammonia. It is not clear if this is a cellular strategy to scavenge free amine-containing compounds. In this example, even though restricticin was not isolated as a pure compound from the heterologous host, formation of acetylated and modified versions of the compound confirmed the role of the BGC, which led to identification of additional CYP51 inhibitors such as lanomycin using genome mining approaches.¹⁶⁶

4.2. Shunt products from pericyclic reactions in biosynthesis

4.2.1. Pericyclase in biosynthesis of 2-pyridones—Pericyclic reactions are among the most powerful chemical transformations to construct multiple regioselective and stereoselective carbon-carbon and carbon-heteroatom bonds via a single transition state.¹⁶⁷ However, it is challenging to control stereoselectivity, regioselectivity and periselectivity of pericyclic reactions without catalysts, especially when competing transition states are accessible. In natural product biosynthesis, these types of selectivity are strictly controlled by pericyclases, a family of enzymes that catalyze pericyclic reactions.¹⁶⁸ In other words, without pericyclases functions, regioisomeric and stereoisomeric pericyclic shunt products are formed in the biosynthetic pathways. During heterologous reconstitution of the biosynthesis of leporin B (**54**) that contains a characteristic dihydropyran core derived from a hetero-Diels–Alder (HDA) reaction, we observed that spontaneous dehydration of the alcohol (**113**) can form both the (*E*)-**114** and (*Z*)-**114** quinone methide intermediates. Subsequent uncatalyzed pericyclic reactions led to a mixture of regio- and stereoisomeric

HDA reaction and intramolecular Diels–Alder (IMDA) reaction shunt products (**115–117**, **119**, **120**) which are not isolated from the original producing strain of leporins (Figure 17A).¹²⁶ The desired product leporin C (**118**) formed via HDA reaction from (*E*)-**114** was only a minor product when no pericyclase was present. Full structural analysis of the shunt products was highly informative, and suggested an enzyme must be present to i) control the stereochemistry of alcohol **113** dehydration to (*E*)-**114**; and 2) to promote the HDA reaction in a stereoselective fashion and to suppress the competing IMDA reactions. Initial bioinformatics analysis of the BGC revealed no obvious gene candidate to catalyze these reactions.

Interestingly, even though no *O*-methylation step is required for leporin B biosynthesis, the BGC contains a putative *O*-MT (LepI). Initial identification and gene knockout studies in the native producing strain showed *lepI* was required for leporin biosynthesis and unidentified shunt products were produced upon deletion.¹⁶⁹ When LepI was coexpressed with other *lep* enzymes in an *A. nidulans* strain that produced the aforementioned mixture of products **115–120**, exclusive production of the desired HDA reaction product **118** was observed without any regioisomeric and stereoisomeric shunt products. Further biochemical characterization of LepI revealed that it is a multifunctional SAM-dependent pericyclase that catalyzes i) stereoselective *anti*-dehydration of **113** to (*E*)-**114**; ii) two pericyclic transformations: undesired IMDA and desired HDA reactions to form **119** and **118**, respectively; and iii) the first enzymatic retro-Claisen rearrangement of **119** to **118**. Co-formation of **118** and **119** was calculated to be thermodynamically unavoidable, since the two compounds bifurcate from an ambimodal *endo* transition state from (*E*)-**114**. LepI therefore catalyzes the additional retro-Claisen rearrangement to convert the shunt product to **118**. This represents a kinetic ‘by-product recycle’ process to overcome thermodynamic limitations and to arrive fully at the desired biosynthetic end product. Ensuing crystal structure analysis showed a modified *O*-MT-like active site, in which a histidine and an arginine were introduced to accelerate the reactions. The LepI-substrate and LepI-product structures also showed the presence of a hydrophilic wall in the active site to suppress formation of the undesired **120** from the *exo* IMDA reaction.¹⁷⁰

This finding echoes Liu and coworkers’ discovery of a bacterial MT-like enzyme, SpnF, in catalyzing the [4+2] cycloaddition reaction during spinosyn biosynthesis.¹⁷¹ Both studies showcased how Nature repurposes a common MT fold to afford precise control of regioselectivity, stereoselectivity and periselectivity of a pericyclic reaction. Heterologous reconstitution of leporin biosynthesis led to a genome mining effort using LepI as a lead to study other pericyclic reactions in 2-pyridone fungal natural products. Heterologous biosynthesis combined with biochemical and computational characterization of the enzymes were used in these studies to arrive at new pericyclases. Examples include the *C*-methyltransferase like enzyme IccD from ilicicolin H (**2**) biosynthetic pathway that catalyzes an inverse electron demand Diels–Alder reaction;³³ *O*-MT-like enzymes PdxI from the pyridoxatin (**121**) biosynthetic pathway that catalyzes an Alder–ene reaction; and a new HDase EpiI from epoxyridone (**122**) biosynthetic pathway (Figure 17B).¹⁷²

4.2.2. Pericyclase in decalin-containing natural products—The decalin motif is found in many natural products produced by bacteria and fungi, and examples of these compounds have been shown throughout this review. The acyclic carbon backbones of compounds containing the decalin core are products of type I PKSs, such as bacterial multimodular and fungal iterative PKSs.¹⁷³ The chemical logic of polyketide decalin formation is through the programming rules of PKSs, either vectorially or iteratively, to strategically recruit (or reject) ER functions in modifications of the polyketide backbones after each chain extension step. This can lead to generation of a diene and dienophile (typically in the form of an α , β -unsaturated olefin) pair interrupted by four contiguous sp^3 carbons that can undergo IMDA cyclization to form a decalin structure. Such chemical logic is also reflected in many biomimetic syntheses of decalin natural products through IMDA reactions.^{174,175} However, as seen in these synthetic efforts, controlling the stereoselectivity of the IMDA reaction is a challenging endeavor, especially for the stereochemical outcomes that must proceed through kinetically unfavorable transition states. Therefore, it has long been speculated that a dedicated pericyclase must be involved in the biosynthesis.

Although several decalin-forming Diels–Alderases have been discovered from bacterial pathways,¹⁷¹ fungal decalin-forming Diels–Alderases were only discovered recently. In 2015, Sato et al. reported the discovery of CghA from the biosynthetic pathway of Sch210972 (**123**) isolated from *Chaetomium globosum*.¹⁷⁶ The structure of **123** contains a *trans*-fused decalin ring system connected to a tetramic acid moiety derived from γ -hydroxymethyl-L-glutamic acid. The *cgh* BGC encodes a PKS-NRPS (CghG), a *trans*-ER (CghC), a lipocalin-like enzyme (CghA), an aldolase (CghB), and a transcription factor (CghD). The aldolase CghB was characterized to catalyze the aldol reaction of two pyruvate molecules to yield γ -hydroxymethyl-L-glutamic acid.¹⁷⁷ The only candidate enzyme remaining in the pathway to catalyze the IMDA reaction is CghA, which is a lipocalin-like protein with unknown function. CghA homologs are found in many fungal BGCs that are connected to products formed via IMDA reactions, including other pyrrolidine-2-one-bearing decalin compounds and cytochalasans such as cytochalasin E (**124**), in which the isoindolone ring system was proposed to be formed by the homologous CcsF.¹⁷⁸ The role of CghA was tested through heterologous expression of CghG, CghC, and CghB with or without CghA in *A. nidulans*.¹⁷⁶ Whereas heterologous expression of CghG, CghC, CghB, and CghA in *A. nidulans* gave the *trans*-decalin **123** exclusively, excluding CghA from the heterologous pathway led to the formation of not only **123** but also the *cis*-decalin diastereomer as the shunt product (**125**) (Scheme 5). **123** and **125** are formed through the *endo* and *exo* transition states, respectively, both of which were computationally predicted to be kinetically accessible. The heterologous expression results therefore indicated CghA is a pericyclase that catalyzes the *endo*-selective IMDA reaction while suppressing the *exo*-IMDA reaction. Recent structural and computational studies of CghA identified key residues in the active site that confer such stereoselectivity through steric interactions with the acyclic intermediate. Mutations to these residues led to inversion of stereoselectivity.¹⁷⁹ At the time of CghA discovery, Osada and coworkers arrived at the same conclusion that Fsa2, a homolog of CghA from the equisetin (**126**) biosynthetic pathway, is a lipocalin-like Diels–Alderase that catalyzes the *endo*-selective decalin-forming Diels–Alder reaction.¹⁸⁰ The

pioneering discoveries of CghA and Fsa2 have led to the identification of other lipocalin-like pericyclases including Eqx3, PvhB,¹⁸¹ MycB,¹²⁵ UcsH,¹⁶⁴ and AspoB.¹⁸²

Depending on molecular orbitals (MO) involved in the reaction, DA reaction can be further classified as normal-electron demand Diels–Alder reactions (NEDDA) and inverse-electron demand Diels–Alder reactions (IEDDA). In the case of NEDDA, the DA reaction happens between an electron-rich dienophile and an electron-deficient diene, while in IEDDA, the reaction takes place between an electron-rich diene and an electron-poor dienophile. Among many known Diels–Alderase characterized so far, the biosynthesis of varicidin A (**127**) involves IEDDA to control the selectivity of IMDA reaction.¹⁸¹ The *pvh* BGC from *Penicillium variable* was targeted for mining because the clustering of a PKS-NRPS (PvhA) and a lipocalin-like enzyme (PvhB). In addition to those two enzymes, the *pvh* cluster also encodes a *trans*-ER (PvhC) that functions collaboratively with PKS-NRPS (PvhA), *N*-methyltransferase (PvhD), and a P450 (PvhE). Coexpression of PvhA with PvhC in *A. nidulans* accumulated three metabolites **128–130** (Figure 18A). **128** is an acyclic tetramate-containing compound that is the product of PKS-NRPS and *trans*-ER. The presence of the diene and dienophile pairs in **128** led to the initial proposal that this is the substrate of the Diels–Alderase. The pair of **129** and **130** were structurally characterized to be *trans*-decalin diastereomers, which are expected to derive from nonenzymatic IMDA of **128**. This further solidified the hypothesis that PvhB may be required to control the stereochemistry to give either **129** or **130** as the desired biosynthetic intermediate. However, further coexpression of PvhB did not change the metabolic profile in *A. nidulans* transformant, suggesting **129** and **130** are in fact shunt products instead of intermediates. Coexpression of the P450 PvhE with PKS-NRPS and *trans*-ER led to the carboxylated compound **131**, which does not undergo nonenzymatic IMDA reaction. Coexpression of PvhB in the above strain exclusively produced the *cis*-decalin **127**, while no other diastereomer can be detected. Therefore, PvhB selectively catalyzes *exo*-cycloaddition of the carboxylated substrate in an IEDDA reaction to give the *cis*-decalin product. DFT calculations on either **128** or **131** showed the activation barrier for reaching the IEDDA reaction transition state is higher than that of the NEDDA reaction. However, upon carboxylation in **131**, the *exo* transition state of IEDDA reaction required to form the *cis*-decalin becomes more kinetically accessible compared to the carboxylated *endo* transition state. Therefore, the biosynthesis of varicidin unveiled an interesting strategy for Nature to access the *cis*-decalin structure: carboxylative deactivation of the nonenzymatic NEDDA reaction followed by enzymatic IEDDA reaction to control the diastereoselectivity.

While shunt products derived from off-pathway oxidation or reduction are undesirable in heterologous reconstitution, in certain cases, such modification may slow down nonenzymatic reactions and can be useful in biochemical assays if the enzymatic recognition moiety in the structure remains unchanged. This was observed in the reconstitution of myceliothermophin A (**132**) biosynthesis using *A. nidulans*.¹²⁵ The structure of **132** contains a *trans*-fused decalin ring system connected to a 3-pyrrolin-2-one moiety derived from reductive release from a PKS-NRPS. A compact three-gene *myc* BGC from *Myceliophthora thermophila* was identified and expressed in *A. nidulans*. Coexpression of PKS-NRPS (MycA) and *trans*-ER (MycC) gave two new metabolites **133** and **134**, both with a 3-

pyrrolin-2-one moiety. **133** is an acyclic polyolefin in the enol form at C₁₈, while **134** is a further oxidized version of **133** with the keto form at C₁₈ (Figure 18B). The oxidation of **133** to **134** is nonenzymatic and involves the generation of hydrogen peroxide. Upon incubation with the lipocalin-like enzyme MycB, no reaction of the enol **133** can be observed, while complete transformation of **134** to the product myceliothermophin E (**135**) was seen. Based on this data, it is proposed that the ketone tautomer of **133**, **136**, is the true substrate of MycB and can undergo cycloaddition to the final product **132**. In the absence of MycB, the inactive enol tautomer **133** can be formed and can undergo oxidation to **134**.

4.2.3. Pericyclase in decahydrofluorene-containing natural products—A lipocalin-like enzyme homologous to CghA, when paired with a PKS-NRPS in a BGC, is a strong indicator of the natural product containing a decalin core, as described above. However, a different role of a CghA homolog was observed in the biosynthesis of pyrrocidines (**137** and **138**).¹⁸³ Pyrrocidines are members of hirsutellone family of natural products which have complex structures and diverse biological activities. Compounds in this family have the highly strained 12- or 13-membered *para*-cyclophane ring D, that is connected through an aryl-ether linkage to a decahydrofluorene core (rings A, B, and C). Because of this highly complex structural feature, hirsutellones have attracted significant interest from synthetic chemists.^{184,185}

Biosynthetically, pyrrocidines are derived from the 3-pyrrolin-2-one intermediate, which is synthesized by a PKS-NRPS in collaboration with a *trans*-ER as suggested by Oikawa.¹⁸⁶ The intermediate is proposed to cyclize into the *para*-cyclophane and form ring C through either electrophilic cyclization or a radical mechanism, connecting the tyrosine phenol and the acyclic polyketide chain, as shown by Nay.¹⁸⁷ A subsequent IMDA reaction is proposed to yield the decahydrofluorene core. Based on the total synthesis of hirsutellone B (**139**), formation of the *cis*-fused A/B ring system in pyrrocidines from the proposed IMDA reaction is highly challenging, which implicates the involvement of an *exo*-specific pericyclase in this pathway. Such retrobiosynthetic analysis led to identification of putative *pyd* cluster which encodes a PKS-NRPS (PydA), *trans*-ER partner (PydC), a lipocalin-like enzyme (PydB), a medium-chain dehydrogenase/reductase (MDR, PydE), an α/β hydrolase (PydG), and three small hypothetical proteins PydX, PydY, and PydZ. Heterologous expression of PydA, PydC, and PydG in *A. nidulans* led to the formation of three shunt products **140–142** (Figure 19). Compound **140** was determined to be a polyolefinic shunt product that is converted from the enolization of the acyclic intermediate (**143**). The shunt product **141** is an air-oxidized shunt product of **140**, while the shunt product **142** contains a cyclohexyl ring that corresponds to ring C in pyrrocidines. To advance the biosynthesis from the acyclic intermediate **143** to pyrrocidines, expression of additional enzymes in *A. nidulans* was performed. While coexpression with the MDR PydE and the putative Diels–Alderase PydB did not yield any new compounds, the additional coexpression of PydX and PydZ in this transformant led to the formation of pyrrocidine D (**144**) that is the *trans* diastereomer of **138**. This suggested the four proteins (PydB, PydE, PydX, and PydZ) may work as a complex to cyclize **143** and form rings D and C in **139**. It is proposed the acyclic chain in **143** is configured in an inverse S-shape (**145**) which enables the phenol to position near C₁₃, and the fully saturated portion of the chain to form a chair-like

conformation aided by the two equatorially substituted methyl groups at C₉ and C₁₁. Upon deprotonation, the phenolate attacks the C₁₃ olefin, which drives the conjugated addition of C₁₂ into the triene at C₇ to form ring C. A hydride acceptor such as the NAD(P)⁺ cofactor in the MDR may be positioned near C₁ to complete the reaction. In the absence of a downstream Diels–Alderase, the resultant product can undergo nonenzymatic cycloaddition through the kinetically accessible *endo* transition state to give the *trans*-fused adduct as the shunt product, which is further modified by a nonenzymatic hydroxylation of C₂, followed by the reduction catalyzed by an ene-reductase in *A. nidulans* to give pyrrocidine D (**144**).

Formation of shunt product **144**, which was not isolated from the pyrrocidine-producing strain, suggested the lipocalin-like enzyme PydB is not the Diels–Alderase responsible for the *exo*-selective cycloaddition required to form **137**. To test if the remaining hypothetical protein PydY is the Diels–Alderase, heterologously coexpression of this enzyme in the strain that produced **144** was performed. The resulting strain indeed produced a new compound that is structurally characterized to be the *cis*-fused decahydrofluorene **146**. PydY, a small lipophilic protein but has no sequence homology to PydB, is therefore the responsible pericyclase in formation of the *cis*-decahydrofluorene. Genome mining of other BGCs in this family showed that homologs of PydY are absent in pathways that make *trans*-decahydrofluorene as seen in **144**.

4.2.4. Nonenzymatic pericyclic reactions—BGCs of natural products formed via a pericyclic reaction do not always contain a pericyclase. In these cases, the lack of stereochemical control leads to formation of multiple diastereomers. For example, nonenzymatic electrocyclization was observed in the biosynthesis of shimalactones (**147** and **148**).¹⁸⁸ Shimalactones possess a characteristic [4.2.0] octadiene structure that is proposed to derive from an 8 π -6 π electrocyclization cascade. The *shm* cluster was found in the producing strain *Emericella varicolor* GF10 by searching for HRPKS with methyltransferase domain but lacking ER domain. The cluster also encodes a FMO (ShmB), a FAD oxidoreductase (ShmF), an integral membrane protein (ShmG), an iron-sulfur protein (ShmD), a transporter (ShmC), and a transcriptional factor (ShmE). Heterologous expression of ShmA in *A. oryzae* led to the production of preshimalactone (**149**) with a conjugated polyene structure. Further expression of ShmA with ShmB resulted in the formation of **147** as the major product and its diastereomer **148** as the minor product. Expression of additional *shm* genes did not change the metabolic profile, suggesting that ShmA and ShmB are sufficient to complete the biosynthesis of shimalactones (Figure 20). These results strongly suggest that ShmB catalyzes the epoxidation of **149** to form preshimalactone epoxide (**150**), which triggers oxabicyclo ring formation and the subsequent nonenzymatic 8 π -6 π electrocyclization to form both **147** and **148**. This proposed mechanism was supported by *in vitro* reaction of **149** with yeast cell-free extract containing ShmB. During the *in vitro* reaction, a new peak with same molecular weight as **147** and **148** was initially observed, followed by the appearance of **147** and **148** peaks in the reaction mixture. This new peak is most likely the epoxide **150**. Based on DFT calculations, after formation of the oxabicyclo ring from **150**, the uncatalyzed 8 π -6 π electrocyclic reaction takes place to give the kinetically favored product **147** as a major product as well as the kinetically disfavored diastereomer **148** as a minor product.

One can conclude that a pericyclase is not involved in shimalactones biosynthesis since both **147** and **148**, diastereomer of shimalactone, have been co-isolated from the same producing strain. **148** thus should not be considered as a shunt product in the biosynthesis of shimalactones. Therefore, discovering a pericyclase that catalyzes electrocyclic reactions remains an objective. It is worth noting that putative “electrocyclases” have been recently reported in bacterial biosynthetic pathways, although these electrocyclases would not be required to control the stereo- and regioselectivity based on the proposed products.¹⁸⁹ We expect that new “electrocyclases” will be discovered from fungal or plant biosynthesis pathways in near future. As described before, such discovery will certainly involve identification of nonenzymatically formed shunt products, which are not formed in the presence of a dedicated “electrocyclase”.

4.3. Shunt products from enzymatic cyclization of polyethers

As seen in the previous example, epoxidation of polyene compounds is a useful strategy to generate complex ring structures. This strategy is most commonly seen in biosynthesis of polyethers derived from polyketides pathways.¹⁹⁰ Baldwin’s rule is widely accepted to explain the relative preference for ring forming reactions according to ring size, position of bond broken, and orbital geometry.¹⁹¹ An example of *anti*-Baldwin’s rule from fungal biosynthesis was observed in aurovertin E (**151**),¹⁹² which has a unique 2,6-dioxabicyclo[3.2.1]octane (DBO) ring system. It is believed this complex bicyclic ring structure is derived from three epoxidation steps and a cascade of regioselective epoxide opening reactions. The gene cluster was found by using HRPKS as a lead. The cluster, *aur*, encodes an HRPKS (AurA), an *O*-MT (AurB), FMO (AurC), a predicted α/β hydrolase (AurD), a putative cyclase (AurE), and an acyltransferase (AurG). The function of each gene was confirmed by gene deletion as well as heterologous reconstitution in yeast. Coexpression of AurABC led to the formation of **152** and other minor metabolites including **153** (Figure 21A). Feeding experiment confirmed that **152** is the only on-pathway intermediate, while other metabolites are shunt products that can be formed by nonenzymatic epoxide opening reactions. Additional expression of AurD with AurABC resulted in the formation of **151** together with a notable decrease in the accumulation of shunt products.

Heterologous expression of the pathway with or without AurD led to the proposed iterative oxidation/cyclization pathway to form **151** from the polyene precursor. First, bis-epoxidation on C₃-C₄ and C₅-C₆ is catalyzed by AurC, which can nonenzymatically cyclize with no regioselective control. In the presence of AurD, regioselective 5-*exo*-tet epoxide ring opening leads to formation of **152**. This is followed by a third epoxidation at C₇-C₈ catalyzed by AurC to form **154**. The more favorable 5-*exo*-tet ring opening reaction as described by the Baldwin’s rule does not take place from **154**. Instead, AurD presumably catalyzes the unfavorable 6-*endo*-tet ring opening to give **151**. It should be noted that Baldwin’s rule is based on nucleophilic attack of the epoxide from an ideal angle. In the presence of an enzyme, the orientation of electrophile and nucleophile can be controlled in the active site leading to deviation of the rule. Structural insights of such deviation have been obtained from Lsd19, which is a polyether epoxide hydrolase that catalyzes disfavored 6-*endo*-tet epoxide ring opening during lasalocid (**155**) biosynthesis (Figure 21B).¹⁹³

4.4. Shunt products from dimeric radical coupling

Control of regio- and stereoselectivity during oxidative coupling reactions is essential in biosynthesis of natural products, especially those derived from dimerization of aryl monomers. In plant biosynthetic pathways, a special class of proteins known as dirigent proteins (from latin dirigere, to guide or align) controls the stereochemistry of radical-based dimerization of phenylpropanoids.¹⁹⁴ For example, the radical-based coupling of *p*-coniferyl alcohol gives a 1:1 mixture of (+) and (–)-pinoresinol in the absence of the dirigent protein.¹⁹⁴ However, in pea and *Arabidopsis thaliana*, dedicated dirigent protein controls stereoselectivity of the dimerization to give predominantly (+)- and (–)-pinoresinol, respectively.^{195,196} A similar mechanism of stereochemical control is exerted by a cotton dirigent protein in the biosynthesis of (+)-gossypol.¹⁹⁷ Recently, a fungal version of dirigent protein that controls atroposelectivity was discovered by Chooi's group through heterologous reconstitution of bisnaphthopyrone viriditoxin biosynthesis¹⁹⁸. The *vdt* BGC was identified using naphtho- α -pyrone structure as a lead. Analysis of the metabolic profile of producing strain revealed that (*M*)-viriditoxin (**156**) is the major enantiomer and (*P*)-viriditoxin (**157**) is the minor product with an approximate 20:1 ratio. This observation suggested that the atroposelectivity of the coupling reaction is controlled by an enzyme. To investigate the pathway, coexpression of NRPKS (VdtA), *O*-methyltransferase (VdtC), Bayer-Villiger Monooxygenase (VdtE), and SDR (VdtF) in *A. nidulans* led to formation of the monomer **158** (Figure 22). Further coexpression VdtACEF with VdtB, which is a multicopper oxidase; and VdtD, a proposed hydrolase, led to the formation of **156** and **157** in 20:1 atropisomeric ratio, same as observed in the producing strain. Removing VdtD from the *A. nidulans* strain, however, the atropisomeric pair was produced at 1:2 ratio. *In vitro* assays using cell-free extracts from *A. nidulans* expressing VdtD and VdtB further supported that VdtD is crucial in controlling the stereoselectivity of biaryl coupling catalyzed by VdtB. This is the first example of a fungal dirigent protein, although the mechanism remains uncharacterized.

5. Conclusion

Heterologous biosynthesis of natural products is a key tool for uncovering cryptic fungal BGCs and assigning functional roles to biosynthetic enzymes. *S. cerevisiae*, *A. nidulans* and *A. oryzae* are the most commonly used hosts, each offering unique advantages. However, it is evident that engineered *Aspergillus* hosts are most preferable since their abilities to correctly splice fungal introns. Genome mining of natural products in heterologous hosts can be driven by structural novelty, biological activity or both. Our review incorporated examples of how to prioritize gene clusters to find new-to-nature structures. Here, we classified these examples as known (types of BGCs) – unknowns (metabolites). However, it is anticipated that the true biosynthetic dark matter will come from the heterologous expression on unknown (no core enzyme)-unknowns (metabolites). With new advances in protein structural prediction (AlphaFold),¹⁹⁹ high-throughput synthetic biology, and small molecular structural elucidation techniques such as MicroED,^{200,201} the goal of fully mapping the fungal secondary metabolome may be realized in the next decade.

Abbreviation used in this manuscript:

4-DMATS	4-dimethylallyl tryptophan synthase
ACP	acyl-carrier protein
ADH	alcohol dehydrogenases
ALS	acetolactate synthase
ATP	adenosine 5'-triphosphate
BGC	biosynthetic gene cluster
Bmt	(4 <i>R</i>)-4-[(<i>E</i>)-2-butenyl]-4-methyl-1-threonine
CAR	carboxylic acid reductase
Cas9	CRISPR-associated protein 9
CBGA	cannabigerolic acid
CRISPR	clustered regularly interspaced short palindromic repeats
Cryo-EM	cryogenic electron microscopy
CPR	cytochrome P450 reductase
DBO	2,6-dioxa-bicyclo[3.2.1]octane
DFT	density functional theory
DHAD	dihydroxyacid dehydratase
DMAPP	dimethylallyl pyrophosphate
DML	dihydromonacolin L
DUF	domain of unknown function
ER	enoyl reductase
FAC	fungus artificial chromosome
FAD	flavin adenine dinucleotide
FMO	flavin-containing monooxygenases
FPP	farnesyl pyrophosphate
GRAS	generally regarded as safe
HDA	hetero-Diels–Alder
HMG-CoA	3-hydroxy-3-methylglutaryl coenzyme A
HP	hypothetical protein

HRPKS	highly-reducing polyketide synthase
IDTC	indole diterpene cyclase
IEDDA	inverse-electron demand Diels–Alder
IMDA	intramolecular Diels–Alder
IPP	isopentenyl pyrophosphate
LC-MS	Liquid Chromatography-Mass Spectrometry
microED	microcrystal electron diffraction
MO	molecular orbital
MT	methyltransferase
NADH	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
NEDDA	normal-electron demand Diels–Alder
NMR	Nuclear Magnetic Resonance
NRPKS	nonreducing polyketide synthase
NRPS	nonribosomal peptide synthetase
OMT	<i>O</i> -methyltransferase
ORF	open reading frame
P450	cytochrome P450
P450_{RE}	ring expansion cytochrome P450
PCR	polymerase chain reaction
PKS	polyketide synthase
PLP	pyridoxal 5'-phosphate
pPant	phosphopantetheinyl
PRPKS	partially-reducing polyketide synthase
PT	prenyltransferase
PTS	post-translational modification
RAL	resorcylic acid lactone
RiPP	ribosomally synthesized and post-translationally modified peptide
RT-PCR	real-time polymerase chain reaction

SDR	short-chain dehydrogenase/reductase
SRE	self-resistance enzyme
TC	terpene cyclase
TF	transcription factor
TS	terpene synthase

Protein domain abbreviation:

KS	ketosynthase
KR	ketoreductase
DH	dehydratase
ER	enoyl reductase
ACP	acyl-carrier protein
AT	acyltransferase
MAT	malonyl-CoA:ACP transacylase
SAT	starter-unit:ACP transacylase
TE	thioesterase
A	adenylation
T	thiolation
PCP	peptidyl-carrier protein
C	condensation
C_T	terminal condensation
R	reduction

6. References

1. Lautru S, Deeth RJ, Bailey LM and Challis GL, *Nat. Chem. Biol.*, 2005, 1, 265–269. [PubMed: 16408055]
2. Zhang JJ, Tang X and Moore BS, *Nat. Prod. Rep.*, 2019, 36, 1313–1332. [PubMed: 31197291]
3. Meng X, Fang Y, Ding M, Zhang Y, Jia K, Li Z, Collemare J and Liu W, *Biotechnol. Adv.*, DOI:10.1016/j.biotechadv.2021.107866.
4. Brandt P, García-Altares M, Nett M, Hertweck C and Hoffmeister D, *Angew. Chemie - Int. Ed.*, 2017, 56, 5937–5941.
5. Dunbar KL, Büttner H, Molloy EM, Dell M, Kumpfmüller J and Hertweck C, *Angew. Chemie - Int. Ed.*, 2018, 57, 14080–14084.
6. Scherlach K and Hertweck C, *Nat. Commun.*, 2021, 12, 1–12. [PubMed: 33397941]

7. Chigumba DN, Mydy LS, de Waal F, Li W, Shafiq K, Wotring JW, Mohamed OG, Mladenovic T, Tripathi A, Sexton JZ, Kautsar S, Medema MH and Kersten RD, *Nat. Chem. Biol.*, 2022, 18, 18–28. [PubMed: 34811516]
8. Brakhage AA, *Nat. Rev. Microbiol.*, 2013, 11, 21–32. [PubMed: 23178386]
9. Wiemann P and Keller NP, *J. Ind. Microbiol. Biotechnol.*, 2014, 41, 301–313. [PubMed: 24146366]
10. Li YF, Tsai KJS, Harvey CJB, Li JJ, Ary BE, Berlew EE, Boehman BL, Findley DM, Friant AG, Gardner CA, Gould MP, Ha JH, Lilley BK, McKinsty EL, Nawal S, Parry RC, Rothchild KW, Silbert SD, Tentilucci MD, Thurston AM, Wai RB, Yoon Y, Aiyar RS, Medema MH, Hillenmeyer ME and Charkoudian LK, *Fungal Genet. Biol.*, 2016, 89, 18–28. [PubMed: 26808821]
11. Robey MT, Caesar LK, Drott MT, Keller NP and Kelleher NL, *Proc. Natl. Acad. Sci. U. S. A.*, DOI:10.1073/pnas.2020230118.
12. Bode HB, Bethe B, Höfs R and Zeeck A, *ChemBioChem*, 2002, 3, 619–627. [PubMed: 12324995]
13. Hewage RT, Aree T, Mahidol C, Ruchirawat S and Kittakoop P, *Phytochemistry*, 2014, 108, 87–94. [PubMed: 25310919]
14. Yuan C, Guo Y-H, Wang H-Y, Ma X-J, Jiang T, Zhao J-L, Zou Z-M and Ding G, *Sci. Rep.*, 2016, 6, 19350. [PubMed: 26839041]
15. Wiemann P, Guo CJ, Palmer JM, Sekonyela R, Wang CCC and Keller NP, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, 110, 17065–17070. [PubMed: 24082142]
16. Bergmann S, Schümann J, Scherlach K, Lange C, Brakhage AA and Hertweck C, *Nat. Chem. Biol.*, 2007, 3, 213–217. [PubMed: 17369821]
17. Ahuja M, Chiang YM, Chang SL, Praseuth MB, Entwistle R, Sanchez JF, Lo HC, Yeh HH, Oakley BR and Wang CCC, *J. Am. Chem. Soc.*, 2012, 134, 8212–8221. [PubMed: 22510154]
18. Chooi YH, Fang J, Liu H, Filler SG, Wang P and Tang Y, *Org. Lett.*, 2013, 15, 780–783. [PubMed: 23368997]
19. Lyu HN, Liu HW, Keller NP and Yin WB, *Nat. Prod. Rep.*, 2020, 37, 6–16. [PubMed: 31033969]
20. Oakley CE, Ahuja M, Sun WW, Entwistle R, Akashi T, Yaegashi J, Guo CJ, Cerqueira GC, Russo Wortman J, Wang CCC, Chiang YM and Oakley BR, *Mol. Microbiol.*, 2017, 103, 347–365. [PubMed: 27775185]
21. Grau MF, Entwistle R, Oakley CE, Wang CCC and Oakley BR, *ACS Chem. Biol.*, 2019, 14, 1643–1651. [PubMed: 31265232]
22. Roux I, Woodcraft C, Hu J, Wolters R, Gilchrist CLM and Chooi YH, *ACS Synth. Biol.*, 2020, 9, 1843–1854. [PubMed: 32526136]
23. Wang C, Hantke V, Cox RJ and Skellam E, *Org. Lett.*, DOI:10.1021/acs.orglett.9b01344.
24. Bian X, Plaza A, Zhang Y and Müller R, *J. Nat. Prod.*, 2012, 75, 1652–1655. [PubMed: 22909174]
25. Olano C, García I, González A, Rodríguez M, Rozas D, Rubio J, Sánchez-Hidalgo M, Braña AF, Méndez C and Salas JA, *Microb. Biotechnol.*, 2014, 7, 242–256. [PubMed: 24593309]
26. Ahmadi MK and Pfeifer BA, *Curr. Opin. Biotechnol.*, 2016, 42, 7–12. [PubMed: 26942861]
27. Myronovskiy M and Luzhetskyy A, *Nat. Prod. Rep.*, 2019, 36, 1281–1294. [PubMed: 31453623]
28. Qu Y, Easson MLAE, Froese J, Simionescu R, Hudlicky T and DeLuca V, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, 112, 6224–6229. [PubMed: 25918424]
29. Lau W and Sattely ES, *Science*, 2015, 349, 1224–1228. [PubMed: 26359402]
30. Nielsen MR, Wollenberg RD, Westphal KR, Sondergaard TE, Wimmer R, Gardiner DM and Sørensen JL, *Fungal Genet. Biol.*, 2019, 132, 103248. [PubMed: 31279974]
31. Shenouda ML, Ambilika M, Skellam E and Cox RJ, *J. Fungi*, DOI:10.3390/jof8040355.
32. Bin Go E, Kim LJ, Nelson HM, Ohashi M and Tang Y, *Org. Lett.*, 2021, 23, 7819–7823. [PubMed: 34581588]
33. Zhang Z, Jamieson CS, Zhao YL, Li D, Ohashi M, Houk KN and Tang Y, *J. Am. Chem. Soc.*, 2019, 141, 5659–5663. [PubMed: 30905148]
34. Asai T, Tsukada K, Ise S, Shirata N, Hashimoto M, Fujii I, Gomi K, Nakagawara K, Kodama EN and Oshima Y, *Nat. Chem.*, 2015, 7, 737–743. [PubMed: 26291946]

35. Tsukada K, Shinki S, Kaneko A, Murakami K, Irie K, Murai M, Miyoshi H, Dan S, Kawaji K, Hayashi H, Kodama EN, Hori A, Salim E, Kuraishi T, Hirata N, Kanda Y and Asai T, *Nat. Commun.*, 2020, 11, 1–12. [PubMed: 31911652]
36. Chankhamjon P, Tsunematsu Y, Ishida-Ito M, Sasa Y, Meyer F, Boettger-Schmidt D, Urbansky B, Menzel KD, Scherlach K, Watanabe K and Hertweck C, *Angew. Chemie - Int. Ed.*, 2016, 55, 11955–11959.
37. Liu M, Ohashi M, Hung YS, Scherlach K, Watanabe K, Hertweck C and Tang Y, *J. Am. Chem. Soc.*, 2021, 143, 7267–7271. [PubMed: 33957045]
38. Zhang JM, Wang HH, Liu X, Hu CH and Zou Y, *J. Am. Chem. Soc.*, 2020, 142, 1957–1965. [PubMed: 31904941]
39. Kakule TB, Jadulco RC, Koch M, Janso JE, Barrows LR and Schmidt EW, *ACS Synth. Biol.*, 2015, 4, 625–633. [PubMed: 25226362]
40. Boecker S, Grätz S, Kerwat D, Adam L, Schirmer D, Richter L, Schütze T, Petras D, Süßmuth RD and Meyer V, *Fungal Biol. Biotechnol.*, 2018, 5, 1–14. [PubMed: 29372063]
41. Nielsen MR, Wollenberg RD, Westphal KR, Sondergaard TE, Wimmer R, Gardiner DM and Sørensen JL, *Fungal Genet. Biol.*, 2019, 132, 103248. [PubMed: 31279974]
42. Pohl C, Polli F, Schütze T, Viggiano A, Mózsik L, Jung S, de Vries M, Bovenberg RAL, Meyer V and Driessen AJM, *Sci. Rep.*, 2020, 10, 1–16. [PubMed: 31913322]
43. Ro DK, Paradise EM, Quellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MCY, Withers ST, Shiba Y, Sarpong R and Keasling JD, *Nature*, 2006, 440, 940–943. [PubMed: 16612385]
44. Brown S, Clastre M, Courdavault V and O'Connor SE, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, 112, 3205–3210. [PubMed: 25675512]
45. Galanie S, Thodey K, Trenchard IJ, Interrante MF and Smolke CD, *Science (80-.)*, 2015, 349, 1095–1100.
46. Awan AR, Blount BA, Bell DJ, Shaw WM, Ho JCH, McKiernan RM and Ellis T, *Nat. Commun.*, 2017, 8, 1–8. [PubMed: 28232747]
47. Luo X, Reiter MA, d'Espaux L, Wong J, Denby CM, Lechner A, Zhang Y, Grzybowski AT, Harth S, Lin W, Lee H, Yu C, Shin J, Deng K, Benites VT, Wang G, Baidoo EEK, Chen Y, Dev I, Petzold CJ and Keasling JD, *Nature*, 2019, 567, 123–126. [PubMed: 30814733]
48. Misa J, Billingsley JM, Niwa K, Yu RK and Tang Y, *ACS Synth. Biol.*, 2022, 11, 1639–1649. [PubMed: 35294193]
49. Lee KKM, Silva NAD and Kealey JT, *Anal. Biochem.*, 2009, 394, 75–80. [PubMed: 19595983]
50. Mootz HD, Schörgendorfer K and Marahiel MA, *FEMS Microbiol. Lett.*, 2002, 213, 51–57. [PubMed: 12127488]
51. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE and Khosla C, *Science (80-.)*, 2001, 291, 1790–1792.
52. Ma SM, Li JWH, Choi JW, Zhou H, Lee KKM, Moorthie VA, Xie X, Kealey JT, Da Silva NA, Vederas JC and Tang Y, *Science (80-.)*, 2009, 326, 589–592.
53. Gao X, Haynes SW, Ames BD, Wang P, Vien LP, Walsh CT and Tang Y, *Nat. Chem. Biol.*, 2012, 8, 823–830. [PubMed: 22902615]
54. Xu W, Cai X, Jung ME and Tang Y, *J. Am. Chem. Soc.*, 2010, 132, 13604–13607. [PubMed: 20828130]
55. Tang MC, Lin HC, Li D, Zou Y, Li J, Xu W, Cacho RA, Hillenmeyer ME, Garg NK and Tang Y, *J. Am. Chem. Soc.*, 2015, 137, 13724–13727. [PubMed: 26469304]
56. Nam W, *Compr. Coord. Chem. II*, 2004, 8, 281–307.
57. Yan Y, Liu Q, Zang X, Yuan S, Bat-Erdene U, Nguyen C, Gan J, Zhou J, Jacobsen SE and Tang Y, *Nature*, 2018, 559, 415–418. [PubMed: 29995859]
58. Harvey CJB, Tang M, Schlecht U, Horecka J, Fischer CR, Lin HC, Li J, Naughton B, Cherry J, Miranda M, Li YF, Chu AM, Hennessy JR, Vandova GA, Inglis D, Aiyar RS, Steinmetz LM, Davis RW, Medema MH, Sattely E, Khosla C, Onge RPS, Tang Y and Hillenmeyer ME, *Sci. Adv.* DOI:10.1126/sciadv.aar5459.
59. Wassarman DA and Steitz JA, *Science (80-.)*, 1992, 257, 1918–1925.

60. Will CL and Lührmann R, *Cold Spring Harb. Perspect. Biol.*, DOI:10.1101/cshperspect.a003707.
61. Zhang X, Guo J, Cheng F and Li S, *Nat. Prod. Rep.*, 2021, 38, 1072–1099. [PubMed: 33710221]
62. Gomi K, Iimura Y and Hara S, *Agric. Biol. Chem.*, 1987, 51, 2549–2555.
63. Jin FJ, Maruyama JI, Juvvadi PR, Arioka M and Kitamoto K, *FEMS Microbiol. Lett.*, 2004, 239, 79–85. [PubMed: 15451104]
64. Heneghan MN, Yakasai AA, Halo LM, Song Z, Bailey AM, Simpson TJ, Cox RJ and Lazarus CM, *ChemBioChem*, 2010, 11, 1508–1512. [PubMed: 20575135]
65. Itoh T, Tokunaga K, Matsuda Y, Fujii I, Abe I, Ebizuka Y and Kushiro T, *Nat. Chem.*, 2010, 2, 858–864. [PubMed: 20861902]
66. Matsuda Y and Abe I, *Nat. Prod. Rep.*, 2016, 33, 26–53. [PubMed: 26497360]
67. Tagami K, Liu C, Minami A, Noike M, Isaka T, Fueki S, Shichijo Y, Toshima H, Gomi K, Dairi T and Oikawa H, *J. Am. Chem. Soc.*, 2013, 135, 1260–1263. [PubMed: 23311903]
68. Liu C, Minami A, Ozaki T, Wu J, Kawagishi H, Maruyama JI and Oikawa H, *J. Am. Chem. Soc.*, 2019, 141, 15519–15523. [PubMed: 31535864]
69. Song Z, Baker W, Marshall JW, Yakasai AA, Khalid RM, Collemare J, Skellam E, Tharreau D, Lebrun MH, Lazarus CM, Bailey AM, Simpson TJ and Cox RJ, *Chem. Sci.*, 2015, 6, 4837–4845. [PubMed: 29142718]
70. Nofiani R, de Mattos-Shiple K, Lebe KE, Han L-C, Iqbal Z, Bailey AM, Willis CL, Simpson TJ and Cox RJ, *Nat. Commun.*, 2018, 9, 3940. [PubMed: 30258052]
71. Nagamine S, Liu C, Nishishita J, Kozaki T, Sogahata K, Sato Y, Minami A, Ozaki T, Schmidt-Dannert C, Ichi Maruyama J and Oikawa H, *Appl. Environ. Microbiol.*, DOI:10.1128/AEM.00409-19.
72. Nayak T, Szewczyk E, Oakley CE, Osmani A, Ukil L, Murray SL, Hynes MJ, Osmani SA and Oakley BR, *Genetics*, 2006, 172, 1557–1566. [PubMed: 16387870]
73. Yee DA and Tang Y, in *Engineering Natural Product Biosynthesis: Methods and Protocols*, ed. Skellam E, Springer US, New York, NY, 2022, pp. 41–52.
74. Chiang YM, Lin TS, Chang SL, Ahn G and Wang CCC, *ACS Synth. Biol.*, 2021, 10, 173–182. [PubMed: 33375785]
75. Somoza AD, Lee K-H, Chiang Y-M, Oakley BR and Wang CCC, *Org. Lett.*, 2012, 14, 972–975. [PubMed: 22296232]
76. Liu N, Hung YS, Gao SS, Hang L, Zou Y, Chooi YH and Tang Y, *Org. Lett.*, 2017, 19, 3560–3563. [PubMed: 28605916]
77. Chiang YM, Ahuja M, Oakley CE, Entwistle R, Asokan A, Zutz C, Wang CCC and Oakley BR, *Angew. Chemie - Int. Ed.*, 2016, 55, 1662–1665.
78. Bok JW, Ye R, Clevenger KD, Mead D, Wagner M, Krerowicz A, Albright JC, Goering AW, Thomas PM, Kelleher NL, Keller NP and Wu CC, *BMC Genomics*, 2015, 16, 1–10. [PubMed: 25553907]
79. Clevenger KD, Bok JW, Ye R, Miley GP, Verdán MH, Velk T, Chen C, Yang KH, Robey MT, Gao P, Lamprecht M, Thomas PM, Islam MN, Palmer JM, Wu CC, Keller NP and Kelleher NL, *Nat. Chem. Biol.*, 2017, 13, 895–901. [PubMed: 28604695]
80. Yan Y, Liu N and Tang Y, *Nat. Prod. Rep.*, 2020, 37, 879–892. [PubMed: 31912842]
81. Almabruk KH, Dinh LK and Philmus B, *ACS Chem. Biol.*, 2018, 13, 1426–1437. [PubMed: 29763292]
82. Lin HC, Chooi YH, Dhingra S, Xu W, Calvo AM and Tang Y, *J. Am. Chem. Soc.*, 2013, 135, 4616–4619. [PubMed: 23488861]
83. Yeh HH, Ahuja M, Chiang YM, Oakley CE, Moore S, Yoon O, Hajovsky H, Bok JW, Keller NP, Wang CCC and Oakley BR, *ACS Chem. Biol.*, 2016, 11, 2275–2284. [PubMed: 27294372]
84. Panter F, Krug D, Baumann S and Müller R, *Chem. Sci.*, 2018, 9, 4898–4908. [PubMed: 29910943]
85. Xie L, Zang X, Cheng W, Zhang Z, Zhou J, Chen M and Tang Y, *J. Am. Chem. Soc.*, 2021, 143, 9575–9584.
86. Biermann F and Helfrich EJN, *mSystems*, 2021, 6, 1–6.

87. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH and Weber T, *Nucleic Acids Res*, 2021, 49, W29–W35. [PubMed: 33978755]
88. Hu Z, Awakawa T, Ma Z and Abe I, *Nat. Commun*, 2019, 10, 1–10. [PubMed: 30602773]
89. Patteson JB, Putz AT, Tao L, Simke WC, Bryant LH, Britt RD and Li B, *Science (80-.)*, 2021, 374, 1005–1009.
90. Lima ST, Fallon TR, Cordoza JL, Chekan JR, Delbaje E, Hopiavuori AR, Alvarenga DO, Wood SM, Luhavaya H, Baumgartner JT, Dörr FA, Etchegaray A, Pinto E, McKinnie SMK, Fiore MF and Moore BS, *J. Am. Chem. Soc*, 2022, 144, 9372–9379. [PubMed: 35583956]
91. Cox RJ, *Org. Biomol. Chem*, 2007, 5, 2010–2026. [PubMed: 17581644]
92. Kasahara K, Miyamoto T, Fujimoto T, Oguri H, Tokiwano T, Oikawa H, Ebizuka Y and Fujii I, *Chembiochem*, 2010, 11, 1245–1252. [PubMed: 20486243]
93. Chooi YH and Tang Y, *J. Org. Chem*, 2012, 77, 9933–9953. [PubMed: 22938194]
94. Crawford JM, Dancy BCR, Hill EA, Udway DW and Townsend CA, *Proc. Natl. Acad. Sci. U. S. A*, 2006, 103, 16728–16733. [PubMed: 17071746]
95. Xie X, Meehan MJ, Xu W, Dorrestein PC and Tang Y, *J. Am. Chem. Soc*, 2009, 131, 8388–8389. [PubMed: 19530726]
96. Xu W, Chooi Y-H, Choi JW, Li S, Vederas JC, Da Silva NA and Tang Y, *Angew. Chem. Int. Ed. Engl*, 2013, 52, 6472–6475. [PubMed: 23653178]
97. Tang MC, Fischer CR, Chari JV, Tan D, Suresh S, Chu A, Miranda M, Smith J, Zhang Z, Garg NK, Onge RPS and Tang Y, *J. Am. Chem. Soc*, 2020, 141, 8198–8206.
98. Ugai T, Minami A, Fujii R, Tanaka M, Oguri H, Gomi K and Oikawa H, *Chem. Commun*, 2015, 51, 1878–1881.
99. Martín I, Peláez F, Harris GH, Curotto JE, Rozdilsky W, Kurtz MB, Giacobbe RA, Bills GF and Cabello MA, *J. Antibiot. (Tokyo)*, 1995, 48, 349–356. [PubMed: 7797434]
100. Li H, Hu J, Wei H, Solomon PS, Stubbs KA and Chooi YH, *Chem. - A Eur. J*, 2019, 25, 15062–15066.
101. Tao H, Mori T, Wei X, Matsuda Y and Abe I, *Angew. Chemie - Int. Ed.*, 2021, 60, 8851–8858.
102. Kennedy J, Auclair K, Kendrew SG, Park C, Vederas JC and Hutchinson CR, *Science (80-.)*, 1999, 284, 1368–1372.
103. Grijseels S, Pohl C, Nielsen JC, Wasil Z, Nygård Y, Nielsen J, Frisvad JC, Nielsen KF, Workman M, Larsen TO, Driessen AJM and Frandsen RJN, *Fungal Biol. Biotechnol*, 2018, 5, 1–17. [PubMed: 29372063]
104. Winter JM, Sato M, Sugimoto S, Chiou G, Garg NK, Tang Y and Watanabe K, *J. Am. Chem. Soc*, 2012, 134, 17900–17903. [PubMed: 23072467]
105. Wang L, Yuan M and Zheng J, *Synth. Syst. Biotechnol*, 2019, 4, 10–15. [PubMed: 30533541]
106. Wang J, Liang J, Chen L, Zhang W, Kong L, Peng C, Su C, Tang Y, Deng Z and Wang Z, *Nat. Commun*, 2021, 12, 1–10. [PubMed: 33397941]
107. Hai Y and Tang Y, *J. Am. Chem. Soc*, 2018, 140, 1271–1274. [PubMed: 29343058]
108. Forseth RR, Amaike S, Schwenk D, Affeldt KJ, Hoffmeister D, Schroeder FC and Keller NP, *Angew. Chemie - Int. Ed.*, 2013, 52, 1590–1594.
109. Baccile JA, Spraker JE, Le HH, Brandenburger E, Gomez C, Bok JW, MacHeleidt J, Brakhage AA, Hoffmeister D, Keller NP and Schroeder FC, *Nat. Chem. Biol*, 2016, 12, 419–424. [PubMed: 27065235]
110. Hai Y, Huang AM and Tang Y, *Proc. Natl. Acad. Sci. U. S. A*, 2019, 116, 10348–10353. [PubMed: 31061132]
111. Cook D, Donzelli BGG, Creamer R, Baucom DL, Gardner DR, Pan J, Moore N, Krasnoff SB, Jaromczyk JW and Schardl CL, *G3 Genes, Genomes, Genet.*, 2017, 7, 1791–1797.
112. Hai Y, Huang A and Tang Y, *J. Nat. Prod*, 2020, 83, 593–600. [PubMed: 32159958]
113. Yun CS, Motoyama T and Osada H, *Nat. Commun*, DOI:10.1038/ncomms9758.
114. Yang X, Feng P, Yin Y, Bushley K, Spatafora JW and Wang C, *MBio*, DOI:10.1128/mBio.01211-18.

115. Gerhards N, Neubauer L, Tudzynski P and Li SM, *Toxins (Basel)*, 2014, 6, 3281–3295. [PubMed: 25513893]
116. Jana N and Nanda S, *New J. Chem*, 2018, 42, 17803–17873.
117. Zhou H, Qiao K, Gao Z, Meehan MJ, Li JWH, Zhao X, Dorrestein PC, Vederas JC and Tang Y, *J. Am. Chem. Soc.*, 2010, 132, 4530–4531. [PubMed: 20222707]
118. Zhou H, Zhan J, Watanabe K, Xie X and Tang Y, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, 105, 6249–6254. [PubMed: 18427109]
119. Al Fahad A, Abood A, Fisch KM, Osipow A, Davison J, Avramovi M, Butts CP, Piel J, Simpson TJ and Cox RJ, *Chem. Sci*, 2014, 5, 523–527. [PubMed: 25580210]
120. Chen M, Liu Q, Gao SS, Young AE, Jacobsen SE and Tang Y, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, 116, 5499–5504. [PubMed: 30842286]
121. Okorafor IC, Chen M and Tang Y, *ACS Synth. Biol*, 2021, 10, 2159–2166. [PubMed: 34415146]
122. Kaneko A, Morishita Y, Tsukada K, Taniguchi T and Asai T, *Org. Biomol. Chem*, 2019, 17, 5239–5243. [PubMed: 31086874]
123. Hashimoto M, Nonaka T and Fujii I, *Nat. Prod. Rep*, 2014, 31, 1306–1317. [PubMed: 25182423]
124. Sims JW, Fillmore JP, Warner DD and Schmidt EW, *Chem. Commun*, 2005, 186–188.
125. Li L, Yu P, Tang MC, Zou Y, Gao SS, Hung YS, Zhao M, Watanabe K, Houk KN and Tang Y, *J. Am. Chem. Soc.*, 2016, 138, 15837–15840. [PubMed: 27960349]
126. Ohashi M, Liu F, Hai Y, Chen M, Cheng Tang M, Yang Z, Sato M, Watanabe K, Houk KN and Tang Y, *Nature*, 2017, 549, 502–506. [PubMed: 28902839]
127. Tokuoka M, Seshime Y, Fujii I, Kitamoto K, Takahashi T and Koyama Y, *Fungal Genet. Biol*, 2008, 45, 1608–1615. [PubMed: 18854220]
128. Sato M, Dander JE, Sato C, Hung YS, Gao SS, Tang MC, Hang L, Winter JM, Garg NK, Watanabe K and Tang Y, *J. Am. Chem. Soc.*, 2017, 139, 5317–5320. [PubMed: 28365998]
129. Christianson DW, *Chem. Rev*, 2017, 117, 11570–11648. [PubMed: 28841019]
130. Du YL and Ryan KS, *Nat. Prod. Rep*, 2019, 36, 430–457. [PubMed: 30183796]
131. Lee C, Chen L, Chiang C, Lai C and Lin H, *Angew. Chemie*, 2019, 131, 18585–18589.
132. Yee DA, Kakule TB, Cheng W, Chen M, Chong CTY, Hai Y, Hang LF, Hung YS, Liu N, Ohashi M, Okorafor IC, Song Y, Tang M, Zhang Z and Tang Y, *J. Am. Chem. Soc.*, 2020, 142, 710–714. [PubMed: 31885262]
133. Tararina MA, Yee DA, Tang Y and Christianson DW, *Biochemistry*, DOI:10.1021/acs.biochem.2c00335.
134. Cheng W, Chen M, Ohashi M and Tang Y, *Angew. Chemie - Int. Ed.*, 2022, 61, 1–5.
135. Yan D, Wang K, Bai S, Liu B, Bai J, Qi X and Hu Y, *J. Am. Chem. Soc.*, 2022, 144, 4269–4276. [PubMed: 35192348]
136. Thomas P, Sundaram H, Krishek BJ, Chazot P, Xie X, Bevan P, Brocchini SJ, Latham CJ, Charlton P, Moore M, Lewis SJ, Thornton DM, Stephenson FA and Smart TG, *J. Pharmacol. Exp. Ther*, 1997, 282, 513–520. [PubMed: 9262310]
137. Davison J, al Fahad A, Cai M, Song Z, Yehia SY, Lazarus CM, Bailey AM, Simpson TJ and Cox RJ, *Proc. Natl. Acad. Sci.*, 2012, 109, 7642–7647. [PubMed: 22508998]
138. Schor R, Schotte C, Wibberg D, Kalinowski J and Cox RJ, *Nat. Commun*, DOI:10.1038/s41467-018-04364-9.
139. Yu F, Okamoto S, Nakasone K, Adachi K, Matsuda S, Harada H, Misawa N and Utsumi R, *Planta*, 2008, 227, 1291–1299. [PubMed: 18273640]
140. Chen Q, Gao J, Jamieson C, Liu J, Ohashi M, Bai J, Yan D, Liu B, Che Y, Wang Y, Houk KN and Hu Y, *J. Am. Chem. Soc.*, 2019, 141, 14052–14056. [PubMed: 31461283]
141. Schotte C, Li L, Wibberg D, Kalinowski J and Cox RJ, *Angew. Chemie - Int. Ed.*, 2020, 59, 23870–23878.
142. Bemis CY, Ungarean CN, Shved AS, Jamieson CS, Hwang T, Lee KS, Houk KN and Sarlah D, *J. Am. Chem. Soc.*, 2021, 143, 6006–6017. [PubMed: 33825475]
143. Kim LJ, Ohashi M, Zhang Z, Tan D, Asay M, Cascio D, Rodriguez JA, Tang Y and Nelson HM, *Nat. Chem. Biol*, 2021, 17, 872–877. [PubMed: 34312563]

144. Singh SB, Liu W, Li X, Chen T, Shafiee A, Card D, Abruzzo G, Flattery A, Gill C, Thompson JR, Rosenbach M, Dreikorn S, Hornak V, Meinz M, Kurtz M, Kelly R and Onishi JC, *ACS Med. Chem. Lett.*, 2012, 3, 814–817. [PubMed: 24900384]
145. Singh SB, Liu W, Li X, Chen T, Shafiee A, Dreikorn S, Hornak V, Meinz M and Onishi JC, *Bioorg. Med. Chem. Lett.*, 2013, 23, 3018–3022. [PubMed: 23562597]
146. Liu L, Tang MC and Tang Y, *J. Am. Chem. Soc.*, 2019, 141, 19538–19541. [PubMed: 31790246]
147. Zhao Z, Ying Y, Hung YS and Tang Y, *J. Nat. Prod.*, 2019, 82, 1029–1033. [PubMed: 30908040]
148. Rank C, Phipps RK, Harris P, Fristrup P, Larsen TO and Gotfredsen CH, *Org. Lett.*, 2008, 10, 401–404. [PubMed: 18179222]
149. Makarova M, Rycek L, Hajicek J, Baidilov D and Hudlicky T, *Angew. Chemie - Int. Ed.*, 2019, 58, 18338–18387.
150. Schaffer JE, Reck MR, Prasad NK and Wencewicz TA, *Nat. Chem. Biol.*, 2017, 13, 737–744. [PubMed: 28504677]
151. Cui Z, Overbay J, Wang X, Liu X, Zhang Y, Bhardwaj M, Lemke A, Wiegmann D, Niro G, Thorson JS, Ducho C and Van Lanen SG, *Nat. Chem. Biol.*, 2020, 16, 904–911. [PubMed: 32483377]
152. Chen M, Liu CT and Tang Y, *J. Am. Chem. Soc.*, 2020, 142, 10506–10515. [PubMed: 32434326]
153. Hai Y, Chen M, Huang A and Tang Y, *J. Am. Chem. Soc.*, 2020, 142, 19668–19677. [PubMed: 33155797]
154. Studt L, Janevska S, Niehaus E-M, Burkhardt I, Arndt B, Sieber CMK, Humpf H-U, Dickschat JS and Tudzynski B, *Environ. Microbiol.*, 2016, 18, 936–956. [PubMed: 26662839]
155. Dai GZ, Han WB, Mei YN, Xu K, Jiao RH, Ge HM and Tan RX, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, 117, 1174–1180. [PubMed: 31882449]
156. Du YL, Singh R, Alkhalaf LM, Kuatsjah E, He HY, Eltis LD and Ryan KS, *Nat. Chem. Biol.*, 2016, 12, 194–199. [PubMed: 26807714]
157. Lo HC, Entwistle R, Guo CJ, Ahuja M, Szewczyk E, Hung JH, Chiang YM, Oakley BR and Wang CCC, *J. Am. Chem. Soc.*, 2012, 134, 4709–4720. [PubMed: 22329759]
158. Gao SS, Zhang T, Garcia-Borràs M, Hung YS, Billingsley JM, Houk KN, Hu Y and Tang Y, *J. Am. Chem. Soc.*, 2018, 140, 6991–6997. [PubMed: 29741874]
159. Gao SS, Duan A, Xu W, Yu P, Hang L, Houk KN and Tang Y, *J. Am. Chem. Soc.*, 2016, 138, 4249–4259. [PubMed: 26978228]
160. Cacho RA, Thuss J, Xu W, Sanichar R, Gao Z, Nguyen A, Vederas JC and Tang Y, *J. Am. Chem. Soc.*, 2015, 137, 15688–15691. [PubMed: 26630357]
161. Cox RJ, *Nat. Prod. Rep.*, DOI:10.1039/d2np00007e.
162. Tang MC, Zou Y, Yee D and Tang Y, *AIChe J.*, 2018, 64, 4182–4186. [PubMed: 31588145]
163. Fujii R, Minami A, Gomi K and Oikawa H, *Tetrahedron Lett.*, 2013, 54, 2999–3002.
164. Li L, Tang MC, Tang S, Gao S, Soliman S, Hang L, Xu W, Ye T, Watanabe K and Tang Y, *J. Am. Chem. Soc.*, 2018, 140, 2067–2071. [PubMed: 29373009]
165. Agatsuma T, Akama T, Nara S, Matsumiya S, Nakai R, Ogawa H, Otaki S, Ikeda S, Saitoh Y and Kanda Y, *Org. Lett.*, 2002, 4, 4387–4390. [PubMed: 12465894]
166. Liu N, Abramyan ED, Cheng W, Perlatti B, Harvey CJB, Bills GF and Tang Y, *J. Am. Chem. Soc.*, 2021, 143, 6043–6047. [PubMed: 33857369]
167. Nicolaou KC, Snyder SA, Montagnon T and Vassilikogiannakis G, *Angew. Chem. Int. Ed. Engl.*, 2002, 41, 1668–1698. [PubMed: 19750686]
168. Jamieson CS, Ohashi M, Liu F, Tang Y and Houk KN, *Nat. Prod. Rep.*, 2019, 36, 698–713. [PubMed: 30311924]
169. Cary JW, Uka V, Han Z, Buyst D, Harris-Coward PY, Ehrlich KC, Wei Q, Bhatnagar D, Dowd PF, Martens SL, Calvo AM, Martins JC, Vanhaecke L, Coenye T, De Saeger S and Di Mavungu JD, *Fungal Genet. Biol.*, 2015, 81, 88–97. [PubMed: 26051490]
170. Cai Y, Hai Y, Ohashi M, Jamieson CS, Garcia-Borràs M, Houk KN, Zhou J and Tang Y, *Nat. Chem.*, 2019, 11, 812–820. [PubMed: 31332284]

171. Kim HJ, Ruszczycky MW, Choi SH, Liu YN and Liu HW, *Nature*, 2011, 473, 109–112. [PubMed: 21544146]
172. Ohashi M, Jamieson CS, Cai Y, Tan D, Kanayama D, Tang MC, Anthony SM, Chari JV, Barber JS, Picazo E, Kakule TB, Cao S, Garg NK, Zhou J, Houk KN and Tang Y, *Nature*, 2020, 586, 64–69. [PubMed: 32999480]
173. Klas K, Tsukamoto S, Sherman DH and Williams RM, *J. Org. Chem*, 2015, 80, 11672–11685. [PubMed: 26495876]
174. Shionozaki N, Yamaguchi T, Kitano H, Tomizawa M, Makino K and Uchiro H, *Tetrahedron Lett*, 2012, 53, 5167–5170.
175. Xu J, Caro-Diaz EJE, Trzoss L and Theodorakis EA, *J. Am. Chem. Soc.*, 2012, 134, 5072–5075. [PubMed: 22390338]
176. Sato M, Yagishita F, Mino T, Uchiyama N, Patel A, Chooi YH, Goda Y, Xu W, Noguchi H, Yamamoto T, Hotta K, Houk KN, Tang Y and Watanabe K, *ChemBioChem*, 2015, 16, 2294–2298. [PubMed: 26360642]
177. Kakule TB, Zhang S, Zhan J and Schmidt EW, *Org. Lett.*, 2015, 17, 2295–2297. [PubMed: 25885659]
178. Qiao K, Chooi Y-H and Tang Y, *Metab. Eng.*, 2011, 13, 723–732. [PubMed: 21983160]
179. Sato M, Kishimoto S, Yokoyama M, Jamieson CS, Narita K, Maeda N, Hara K, Hashimoto H, Tsunematsu Y, Houk KN, Tang Y and Watanabe K, *Nat. Catal.*, 2021, 4, 223–232. [PubMed: 35873532]
180. Kato N, Nogawa T, Hirota H, Jang JH, Takahashi S, Ahn JS and Osada H, *Biochem. Biophys. Res. Commun.*, 2015, 460, 210–215. [PubMed: 25770422]
181. Tan D, Jamieson CS, Ohashi M, Tang MC, Houk KN and Tang Y, *J. Am. Chem. Soc.*, 2019, 141, 769–773. [PubMed: 30609896]
182. Zhang JM, Liu X, Wei Q, Ma C, Li D and Zou Y, *Nat. Commun.*, DOI:10.1038/s41467-021-27931-z.
183. Ohashi M, Kakule TB, Tang MC, Jamieson CS, Liu M, Zhao YL, Houk KN and Tang Y, *J. Am. Chem. Soc.*, 2021, 143, 5605–5609. [PubMed: 33834778]
184. Uchiro H, Kato R, Arai Y, Hasegawa M and Kobayakawa Y, *Org. Lett.*, 2011, 13, 6268–6271. [PubMed: 22040033]
185. Sugata H, Inagaki K, Ode T, Hayakawa T, Karoji Y, Baba M, Kato R, Hasegawa D, Tsubogo T and Uchiro H, *Chem. - An Asian J.*, 2017, 12, 628–632.
186. Oikawa H, *J. Org. Chem.*, 2003, 68, 3552–3557. [PubMed: 12713359]
187. Ear A, Amand S, Blanchard F, Blond A, Dubost L, Buisson D and Nay B, *Org. Biomol. Chem.*, 2015, 13, 3662–3666. [PubMed: 25675395]
188. Fujii I, Hashimoto M, Konishi K, Unezawa A, Sakuraba H, Suzuki K, Tsushima H, Iwasaki M, Yoshida S, Kudo A, Fujita R, Hichiwa A, Saito K, Asano T, Ishikawa J, Wakana D, Goda Y, Watanabe A, Watanabe M, Masumoto Y, Kanazawa J, Sato H and Uchiyama M, *Angew. Chemie - Int. Ed.*, 2020, 59, 8464–8470.
189. Zhang J, Yuzawa S, Thong WL, Shinada T, Nishiyama M and Kuzuyama T, *J. Am. Chem. Soc.*, 2021, 143, 2962–2969. [PubMed: 33576619]
190. Gallimore AR, *Nat. Prod. Rep.*, 2009, 26, 266–280. [PubMed: 19177224]
191. Baldwin BJE, *J.C.S. Chem. Comm.*, 1976, 734–736.
192. Mao XM, Zhan ZJ, Grayson MN, Tang MC, Xu W, Li YQ, Yin WB, Lin HC, Chooi YH, Houk KN and Tang Y, *J. Am. Chem. Soc.*, 2015, 137, 11904–11907. [PubMed: 26340065]
193. Hotta K, Chen X, Paton RS, Minami A, Li H, Swaminathan K, Mathews II, Watanabe K, Oikawa H, Houk KN and Kim CY, *Nature*, 2012, 483, 355–358. [PubMed: 22388816]
194. Davin LB, Bin Wang H, Crowell AL, Bedgar DL, Martin DM, Sarkanen S and Lewis NG, *Science (80-.)*, 1997, 275, 362–366.
195. Seneviratne HK, Dalisay DS, Kim K-W, Moinuddin SGA, Yang H, Hartshorn CM, Davin LB and Lewis NG, *Phytochemistry*, 2015, 113, 140–148. [PubMed: 25457488]
196. Kim K-W, Moinuddin SGA, Atwell KM, Costa MA, Davin LB and Lewis NG, *J. Biol. Chem.*, 2012, 287, 33957–33972. [PubMed: 22854967]

197. Liu J, Stipanovic RD, Bell AA, Puckhaber LS and Magill CW, *Phytochemistry*, 2008, 69, 3038–3042. [PubMed: 18639908]
198. Hu J, Li H and Chooi YH, *J. Am. Chem. Soc.*, 2020, 141, 8068–8072.
199. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P and Hassabis D, *Nature*, 2021, 596, 583–589. [PubMed: 34265844]
200. Jones CG, Martynowycz MW, Hattne J, Fulton TJ, Stoltz BM, Rodriguez JA, Nelson HM and Gonen T, *ACS Cent. Sci.*, 2018, 4, 1587–1592. [PubMed: 30555912]
201. Gruene T, Wennmacher JTC, Zaubitzer C, Holstein JJ, Heidler J, Fecteau-Lefebvre A, De Carlo S, Müller E, Goldie KN, Regeni I, Li T, Santiso-Quinones G, Steinfeld G, Handschin S, van Genderen E, van Bokhoven JA, Clever GH and Pantelic R, *Angew. Chemie - Int. Ed.*, 2018, 57, 16313–16317.

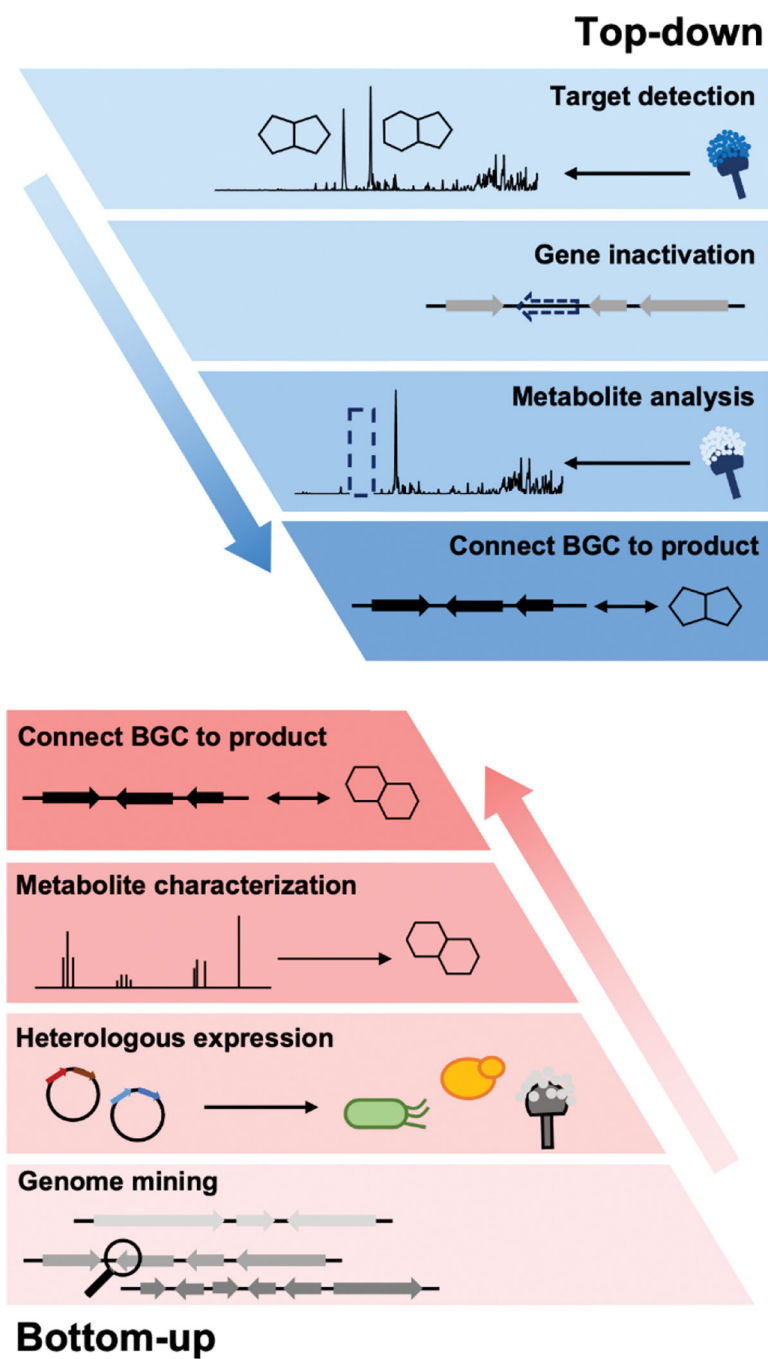


Figure 1.
Top-down and Bottom-up approaches in studying biosynthesis

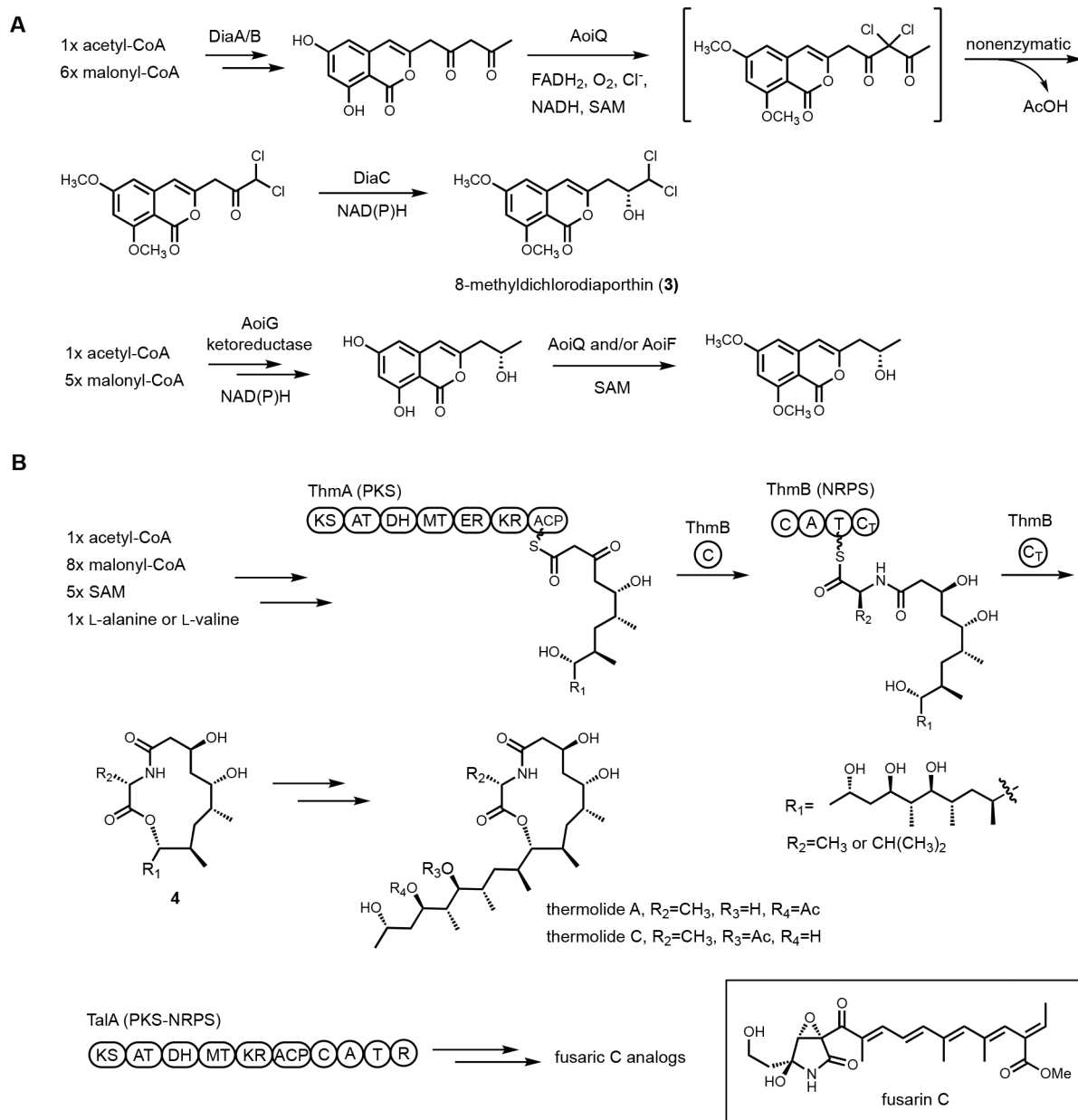


Figure 2. Cluster reassignment using heterologous expression of BGCs. **(A)** Revised biosynthetic pathway of dichlorodiaporthins. **(B)** Revised biosynthetic pathway of thermolides.

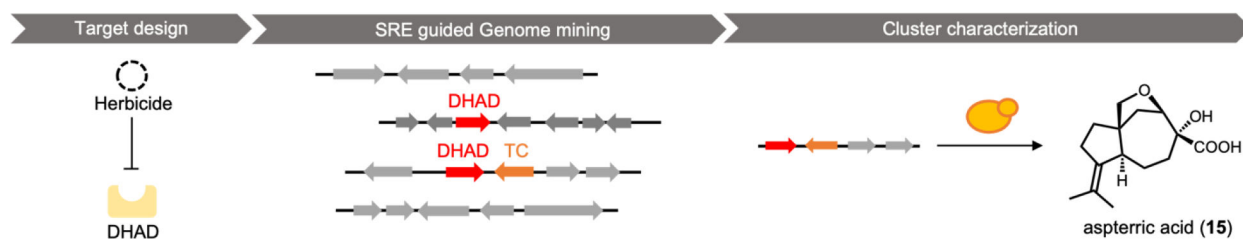


Figure 3.

Example of Self-Resistance Enzyme (SRE)-guided genome mining. With DHAD as a target, the cluster of aspterric acid was identified through SRE-guided genome mining. The product of the cluster was revealed by heterologous expression in yeast.

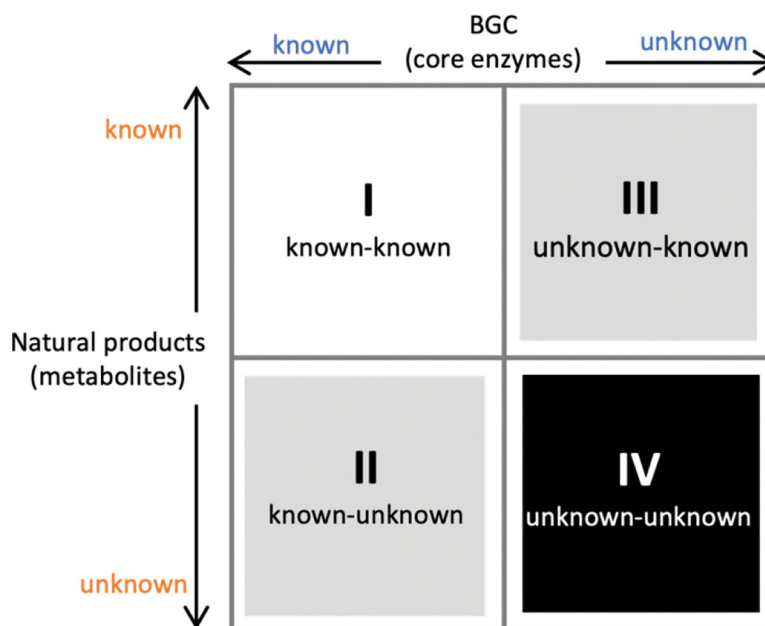


Figure 4. Natural product classification using a quadrant system. X-axis refers to known or unknown gene clusters based on core enzyme prediction. Y-axis refers to known or unknown metabolites. This review mainly focuses on the second quadrant.

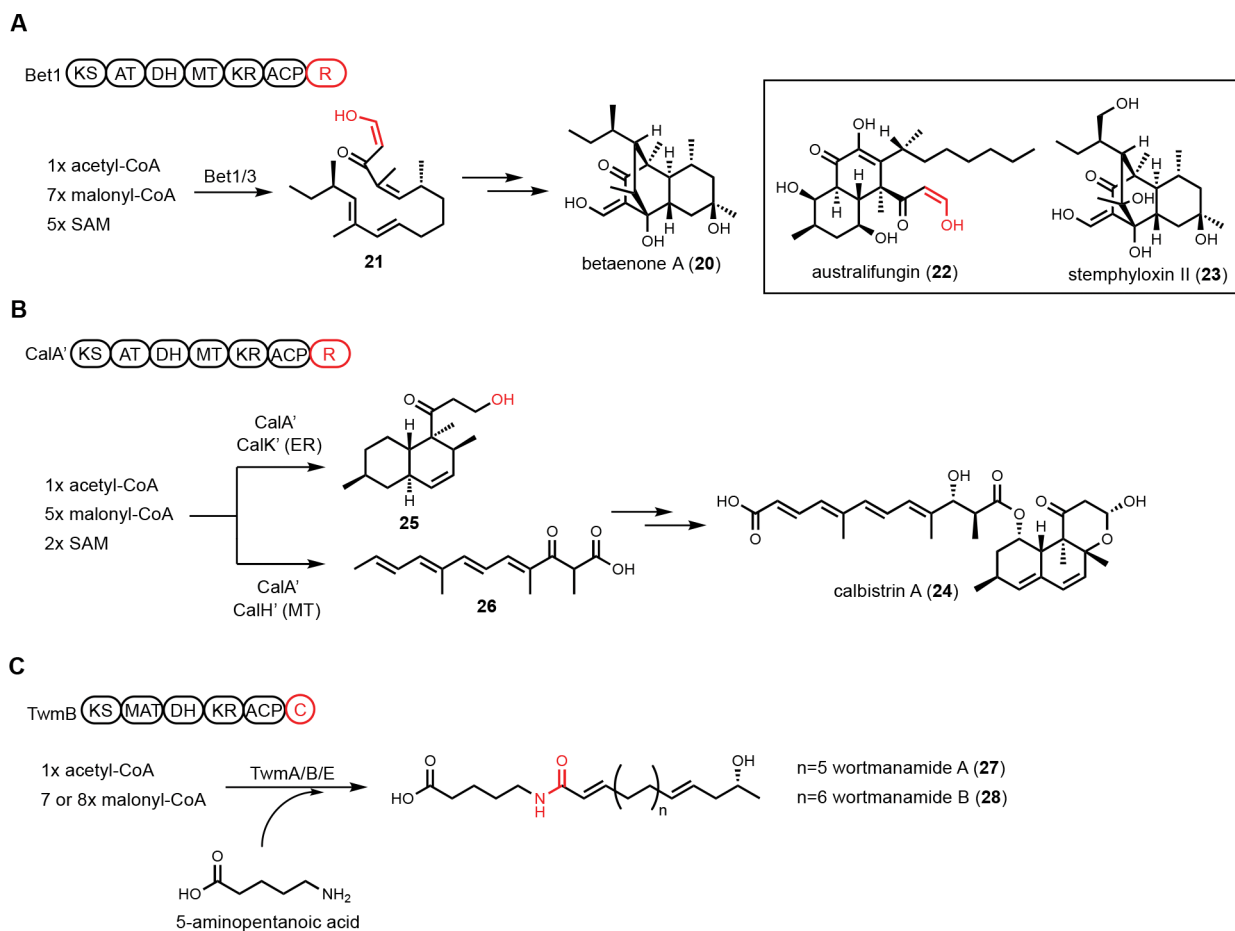


Figure 5. Unusual HRPKS domain arrangement. **(A)** PKS-R in betaenone biosynthesis. Terminal R domain catalyzes reductive product release. The enol form of aldehyde is highlighted in red. **(B)** PKS-R in calbistrin A biosynthesis. CalA' can collaborate with different tailoring enzymes to generate two different polyketide scaffolds. Terminal R domain catalyzes consecutive two-electron reduction to release the decalin product as an alcohol, which is highlighted in red. **(C)** PKS-C in wortmanamide biosynthesis. The C domain fused with PKS is able to catalyze long chain *N*-acyl amide formation.

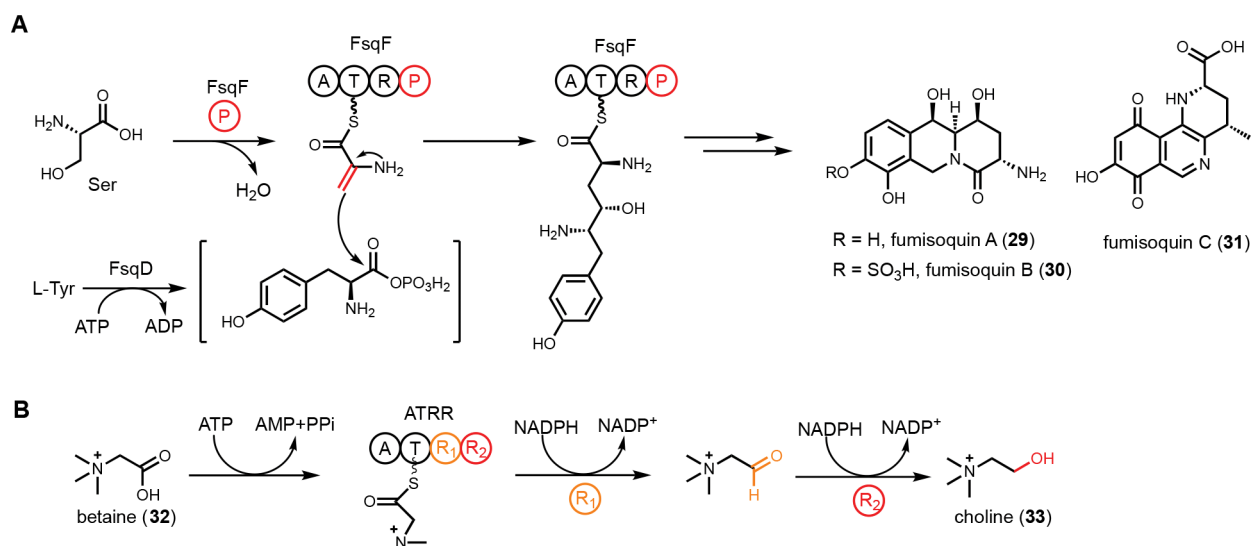


Figure 6.

Noncanonical single module NRPS-like enzyme. **(A)** FsqF (A-T-R-PLP) in fumisoquin biosynthesis. PLP domain is fused with NRPS, catalyzing dehydration of serine and forming a new C-C bond with FsqD-activated tyrosine. **(B)** ATRR (A-T-R₁-R₂) in choline biosynthesis. Two consecutive R domains catalyze sequential reduction to convert betaine to choline.

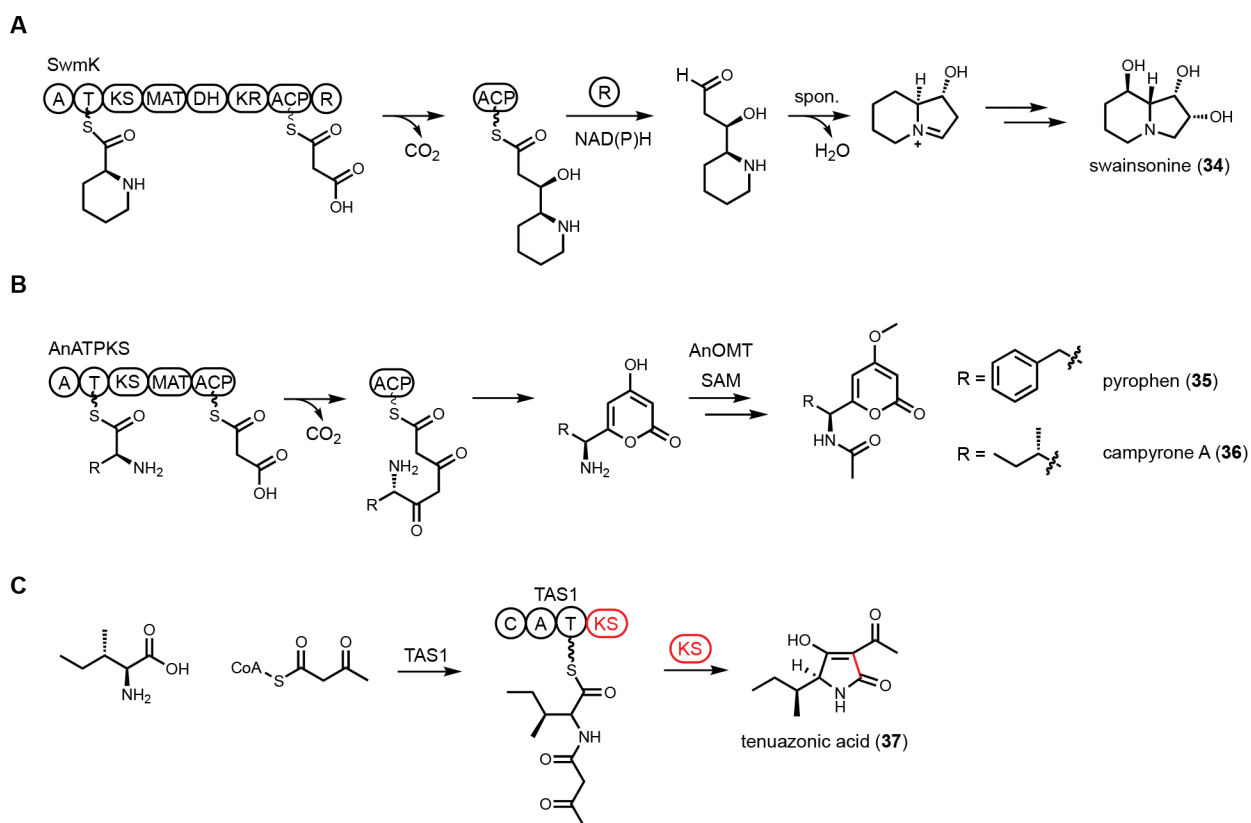


Figure 7. NRPS-PKS hybrid enzymes. **(A)** Proposed swainsonine biosynthetic pathway. SwmK catalyzes the formation of indolizidine core. **(B)** Proposed pyrophen and campyrone biosynthetic pathway. AnATPKS is involved in the formation of amino acid-containing α -pyrone natural products. **(C)** Proposed tenuazonic acid biosynthetic pathway. A KS domain is fused with NRPS to catalyze cyclization and product release.

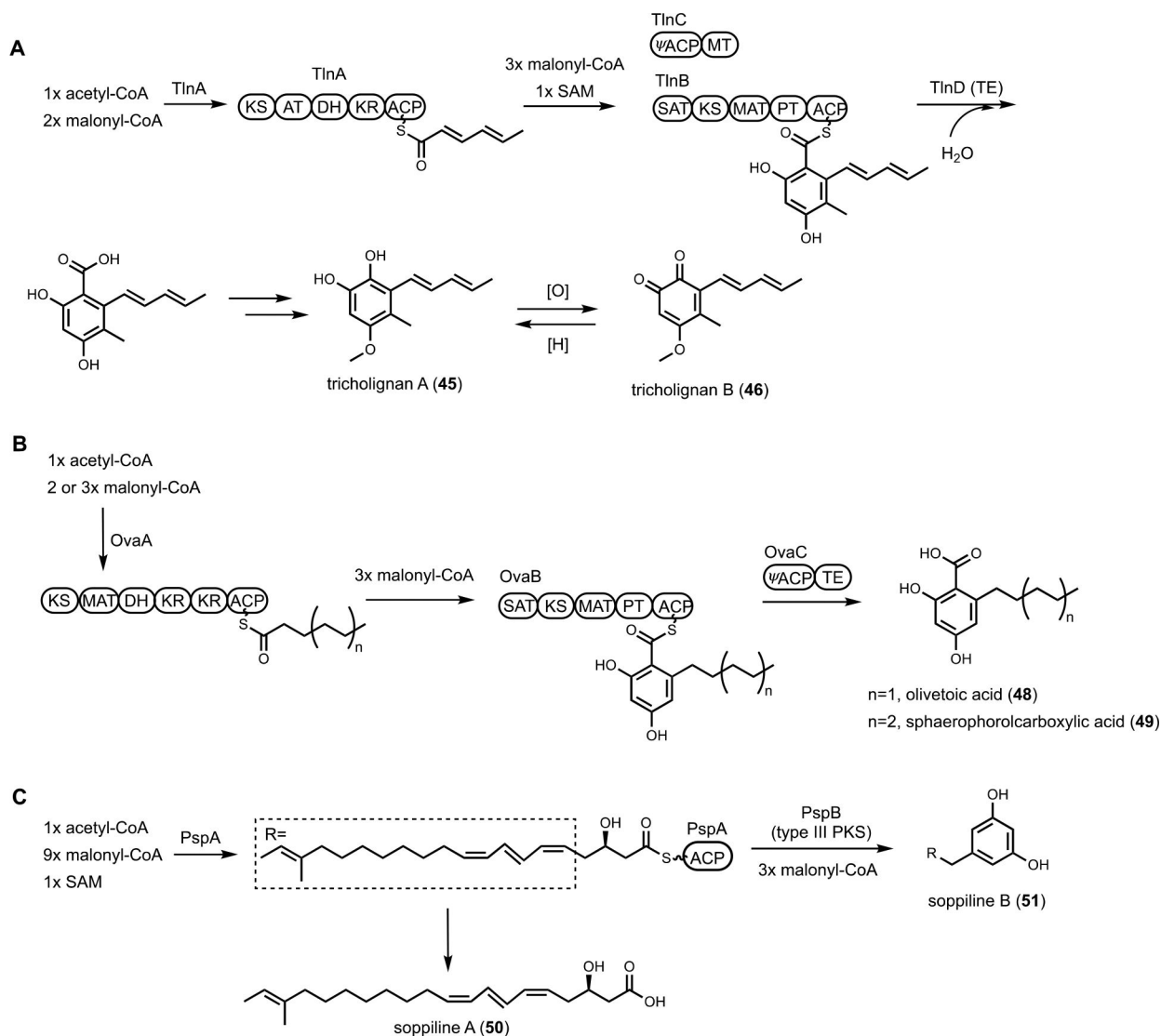
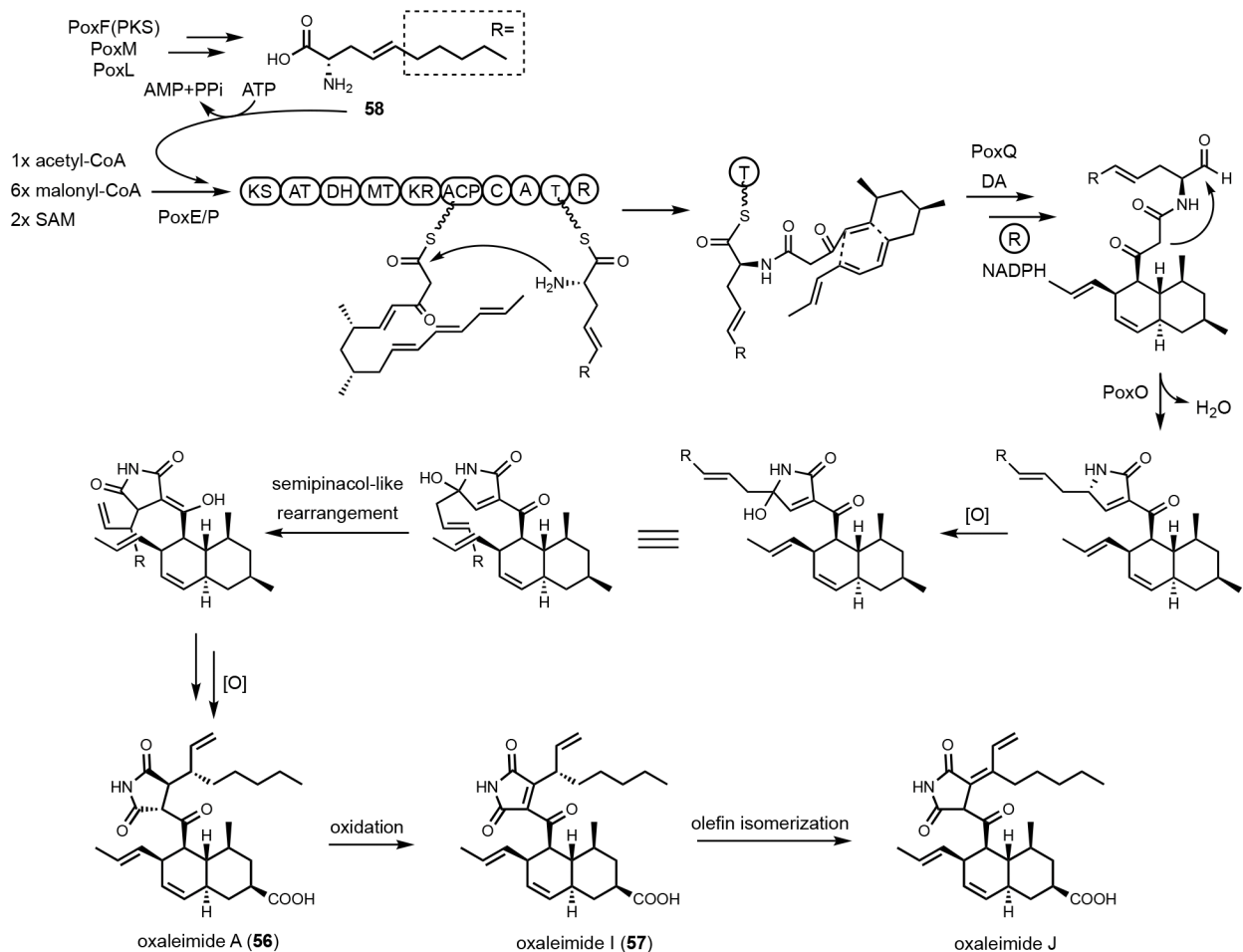
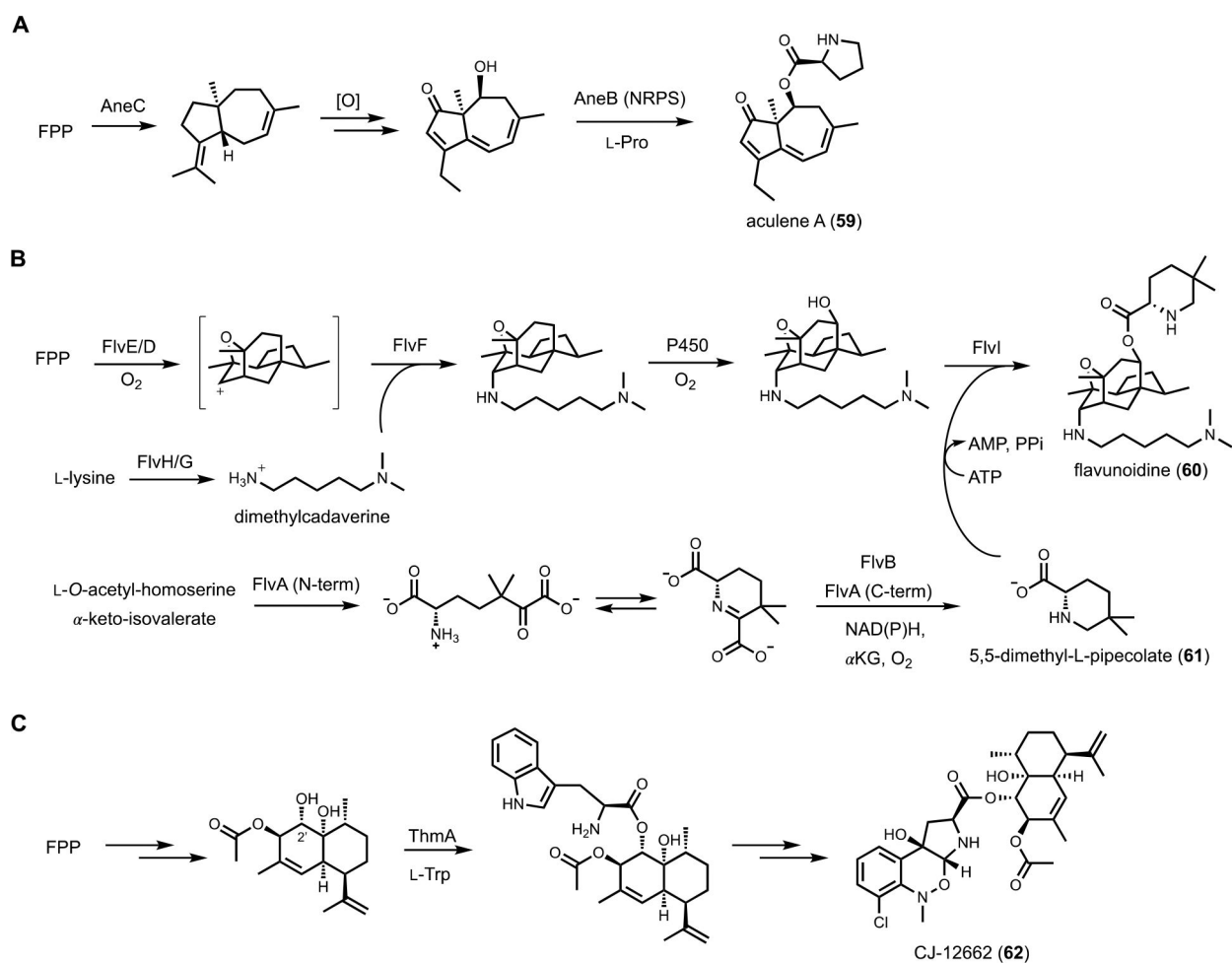


Figure 8. Pathways in which multiple PKSs collaborate. **(A)** Biosynthetic pathway of tricholignan A and B. The pathway requires TlnC, which is a didomain protein with a ψ ACP fused with MT. **(B)** A fungal biosynthetic pathway to olivetolic acid. The cluster contains a didomain enzyme with an ψ ACP fused with TE. **(C)** Proposed pathway of soppilone. A type III PKS collaborates with HRPKS in chain elongation and product cyclization.

**Figure 9.**

Proposed biosynthetic pathway of oxaleimides. The biosynthesis requires HRPKS (PoxF) with PoxM and PoxL to generate a nonproteinogenic amino acid **58**, which is incorporated by PKS-NRPKS (PoxE) assembly line. Knoevenagel condensation followed by hydroxylation and a semipinacol-like rearrangement complete the formation of the substituted succinimide.

**Figure 10.**

Examples of natural products produced from BGCs that encode both terpene cyclase and NRPS. FPP: farnesyl diphosphate. **(A)** Simplified aculene biosynthesis pathway. The proline moiety is incorporated into terpene scaffold via a single module NRPS AneB. **(B)** Proposed flavunoidine biosynthetic pathway. Two unique amino acid building blocks are highlighted in the pathway. *N,N*-dimethylcadaverine is synthesized from L-lysine, which then connected to the tetracyclic core via an axial C-N bond. The other substituent, 5,5-dimethyl-L-pipecolate is formed by PLP-dependent enzyme FivA. The connection of 5,5-dimethyl-L-pipecolate to the tetracyclic structure is catalyzed by a single module NRPS FivI. **(C)** Simplified CJ12662 biosynthesis pathway. L-tryptophan is esterified with the terpene scaffold in an ATP-dependent step catalyzed by ThmA.

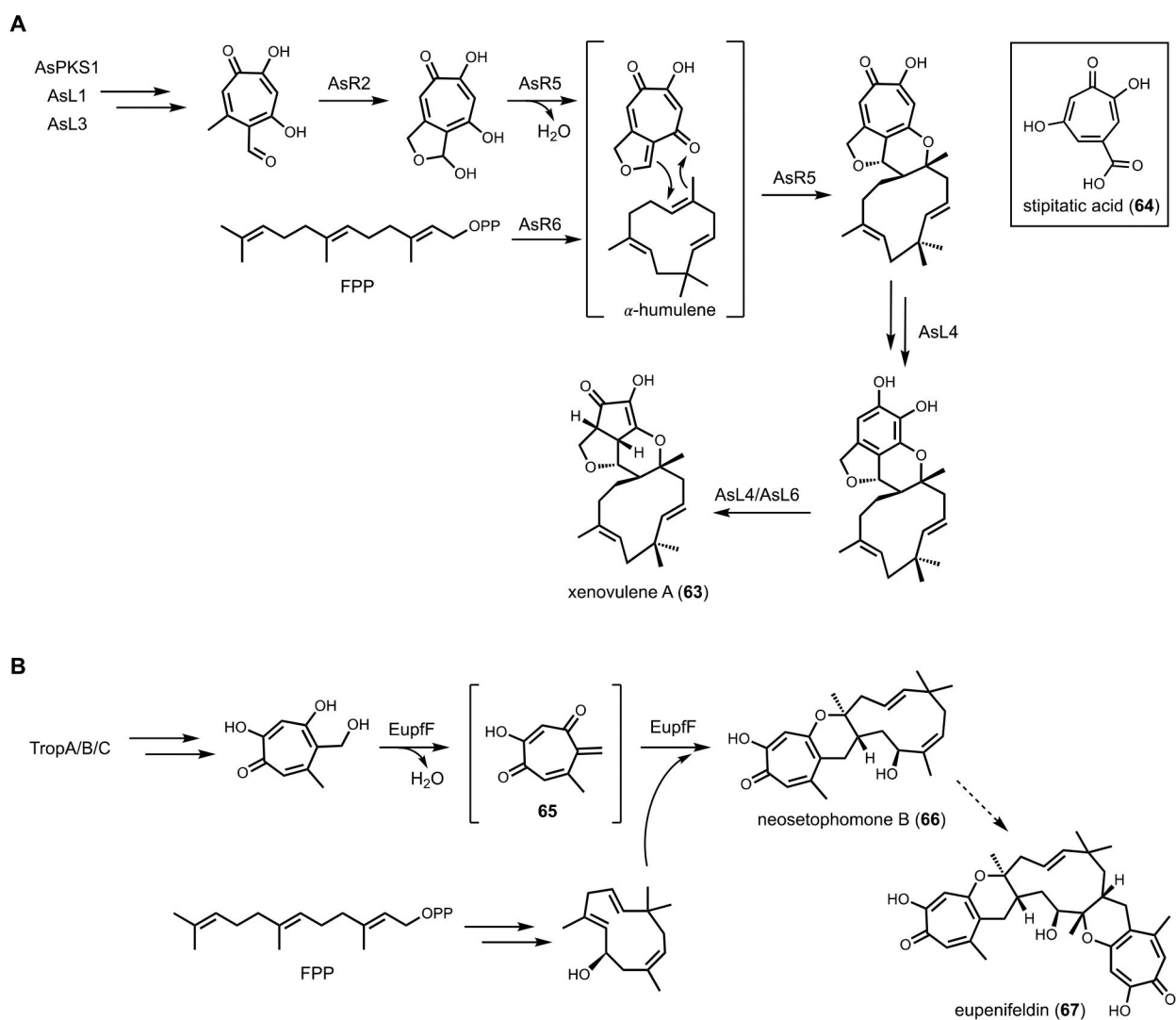
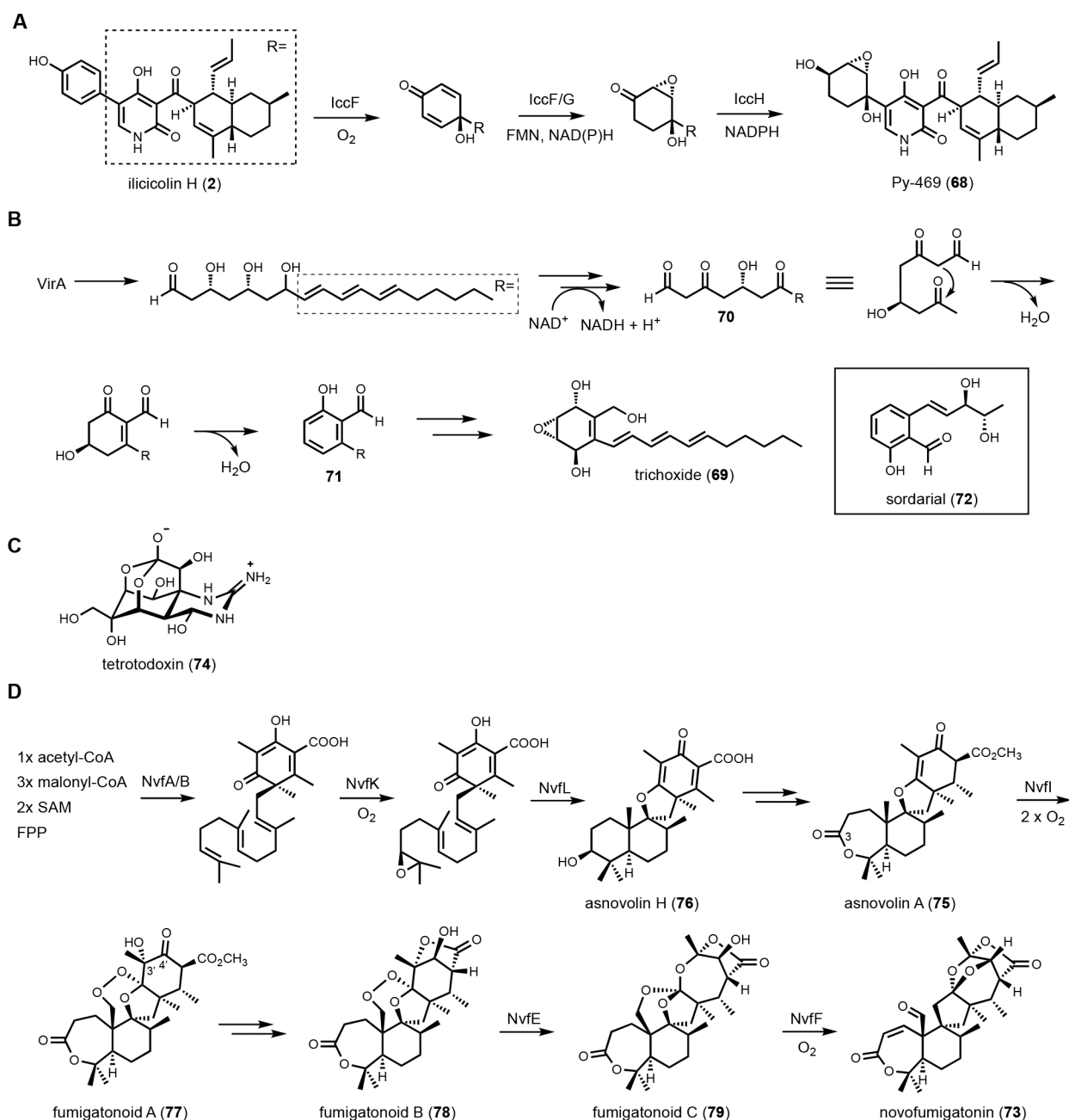
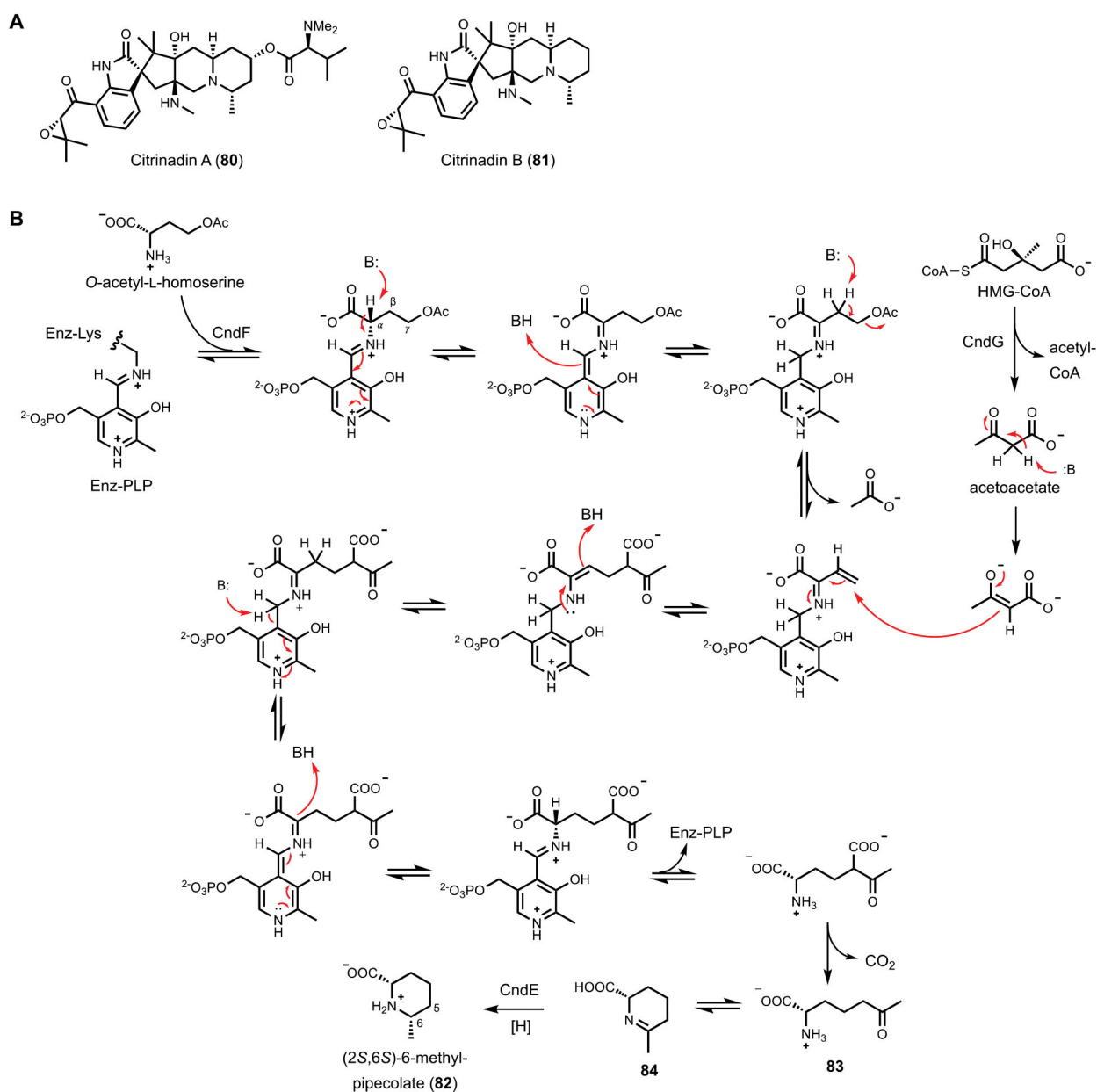


Figure 11.

Cryptic terpene cyclase functions with NRPKS. **(A)** Proposed biosynthetic pathway of xenovulene A. AsR6 was confirmed as a new class of terpene cyclase, catalyzing the formation of humulene. **(B)** Proposed biosynthetic pathway of eupenifeldin. EupfF was characterized as a hetero Diels–Alderase, catalyzing the cycloaddition between the quinone methide **65** and sesquiterpene.

**Figure 12.**

Post-core enzyme modifications by redox enzymes. **(A)** Biosynthetic pathway of Py-469 highlights three consecutive redox enzymes in modifying benzene ring to 2,3-epoxy-1,4-cyclohexane diol. The relative stereochemistry was determined by MicroED. **(B)** Multitude of redox enzymes in the biosynthesis of trichoxide reveals a unique strategy to synthesize aromatic polyketide structure using HRPKS. A similar pathway was also found in sordarial biosynthesis. **(C)** Structure of tetrodotoxin, which is an orthoester containing natural product. **(D)** Simplified biosynthetic pathway of novofumigatonin. The structure is highly oxygenated with a series of oxygen-handling enzymes, in particular the formation of orthoester functional group by two non-heme iron enzymes, NvfE and NvfF.

**Figure 13.**

(A) Structure of citrinadins. (B) Biosynthetic pathway of forming (2*S*,3*S*)-6-methylpipercolate. The mechanism of γ -substitution catalyzed by CndF was proposed. O-acetyl-L-homoserine was first bound to PLP as the external aldimine. After eliminating acetate, acetoacetate generated from HMG-CoA by HMG-CoA lyase then attacks the γ position to form new C-C bond.

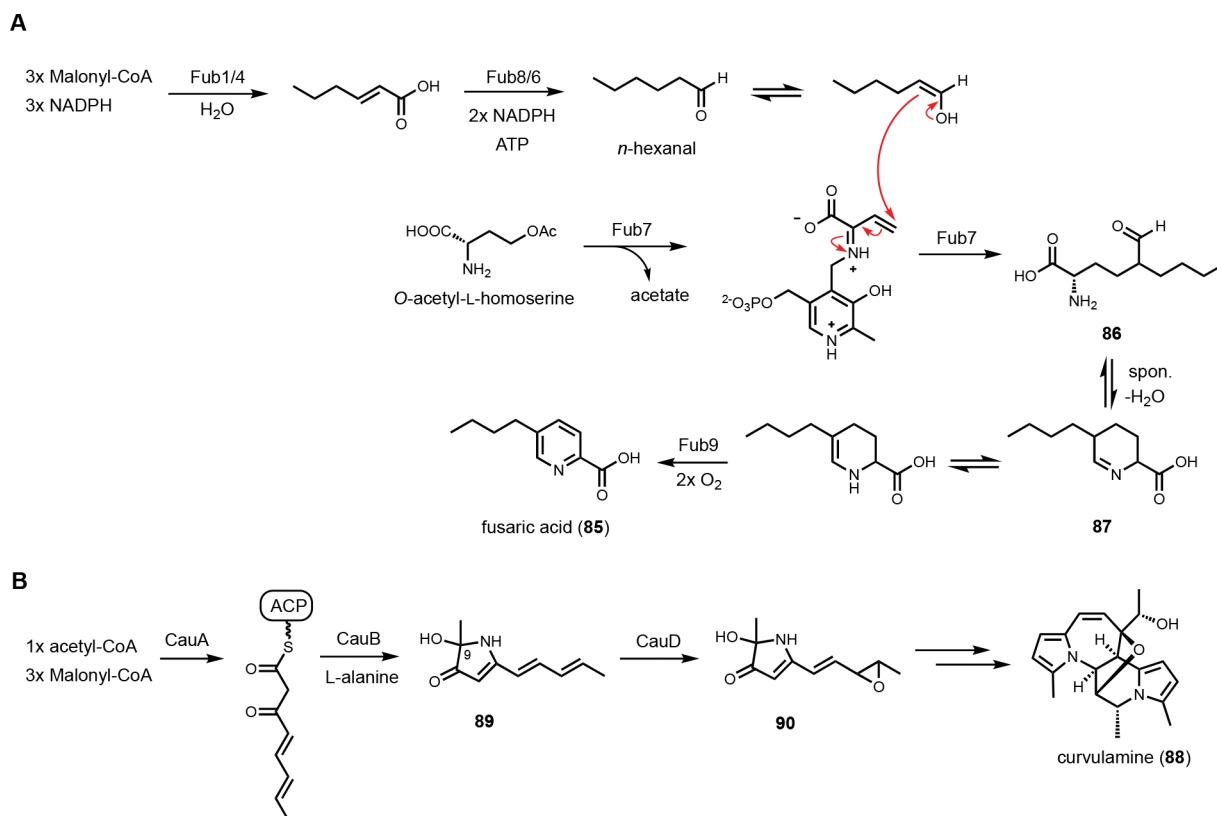


Figure 14.

(A) Proposed biosynthetic pathway of fusaric acid. Fub7 is a PLP-dependent enzyme that catalyzes γ -replacement C-C bond formation. The mechanism is similar to CndF. (B) Simplified biosynthetic pathway of curvulamine, highlighting a bifunctional PLP-dependent enzyme, CauB, that can catalyze Claisen condensation to release PKS product as well as hydroxylation on the substrate.

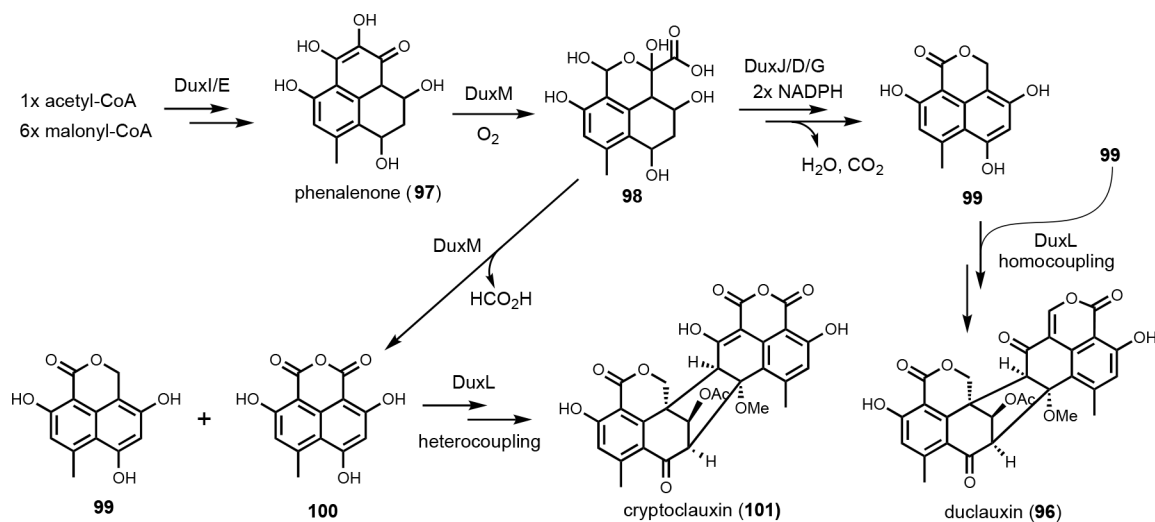


Figure 15.

Proposed pathway of duclauxins. DuxM is the key enzyme to start the redox modifications, leading to the production of two different monomers **99** and **100**. DuxL then catalyzes coupling reaction to give duclauxin or cryptoclauxin.

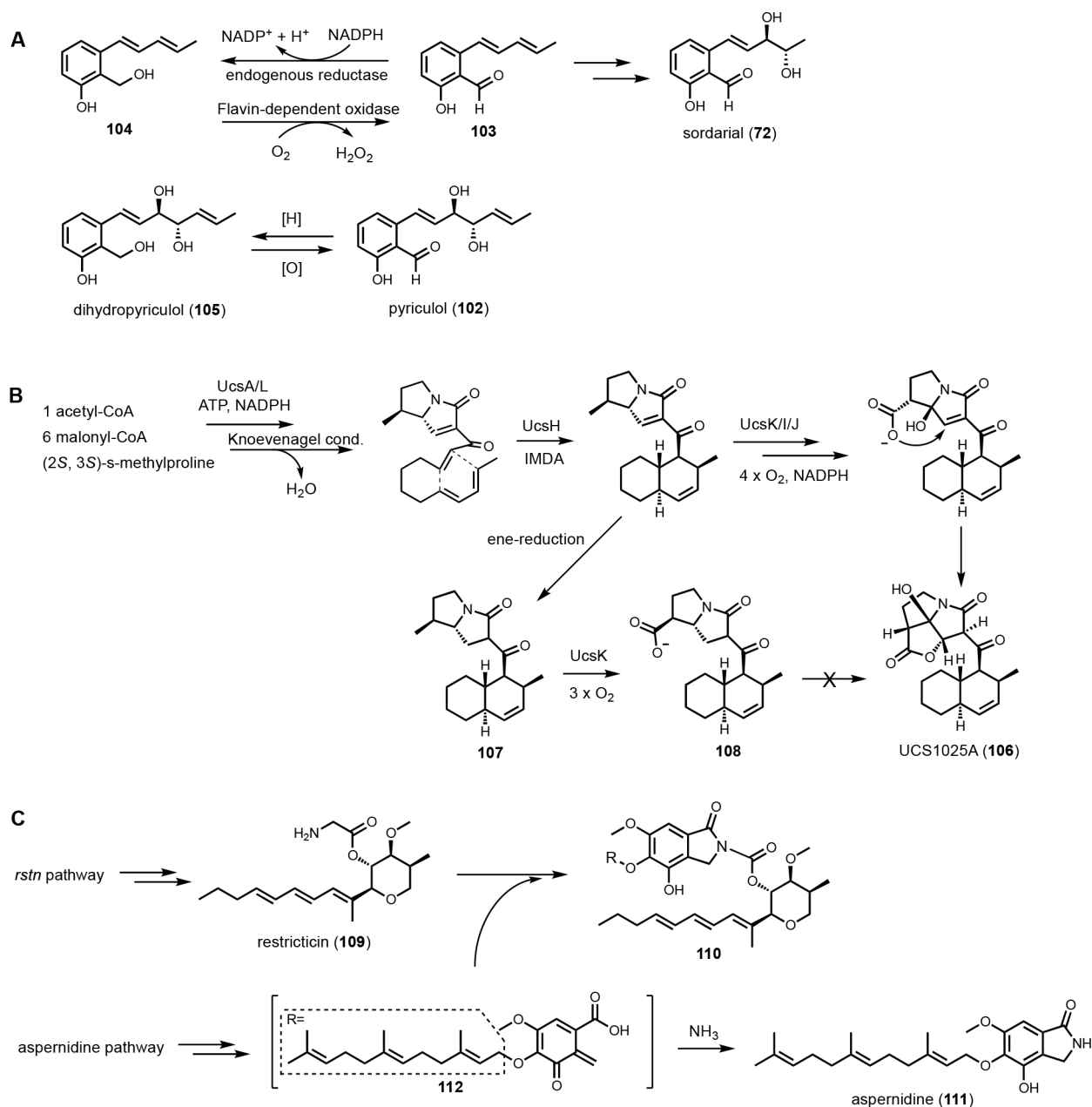


Figure 16. Shunt products formed by cellular crosstalk. **(A)** Two examples of endogenous NADPH-dependent reduction convert aldehydes to alcohols. The shunt products can be reoxidized by flavin-dependent oxidase to be further modified in the biosynthetic pathway of salicylaldehyde products. **(B)** Biosynthesis of UCS1025A as an example of ene reduction hampering heterologous reconstitution. **(C)** Proposed cellular crosstalk between restricticin and *A. nidulans* endogenous metabolite aspernidine.

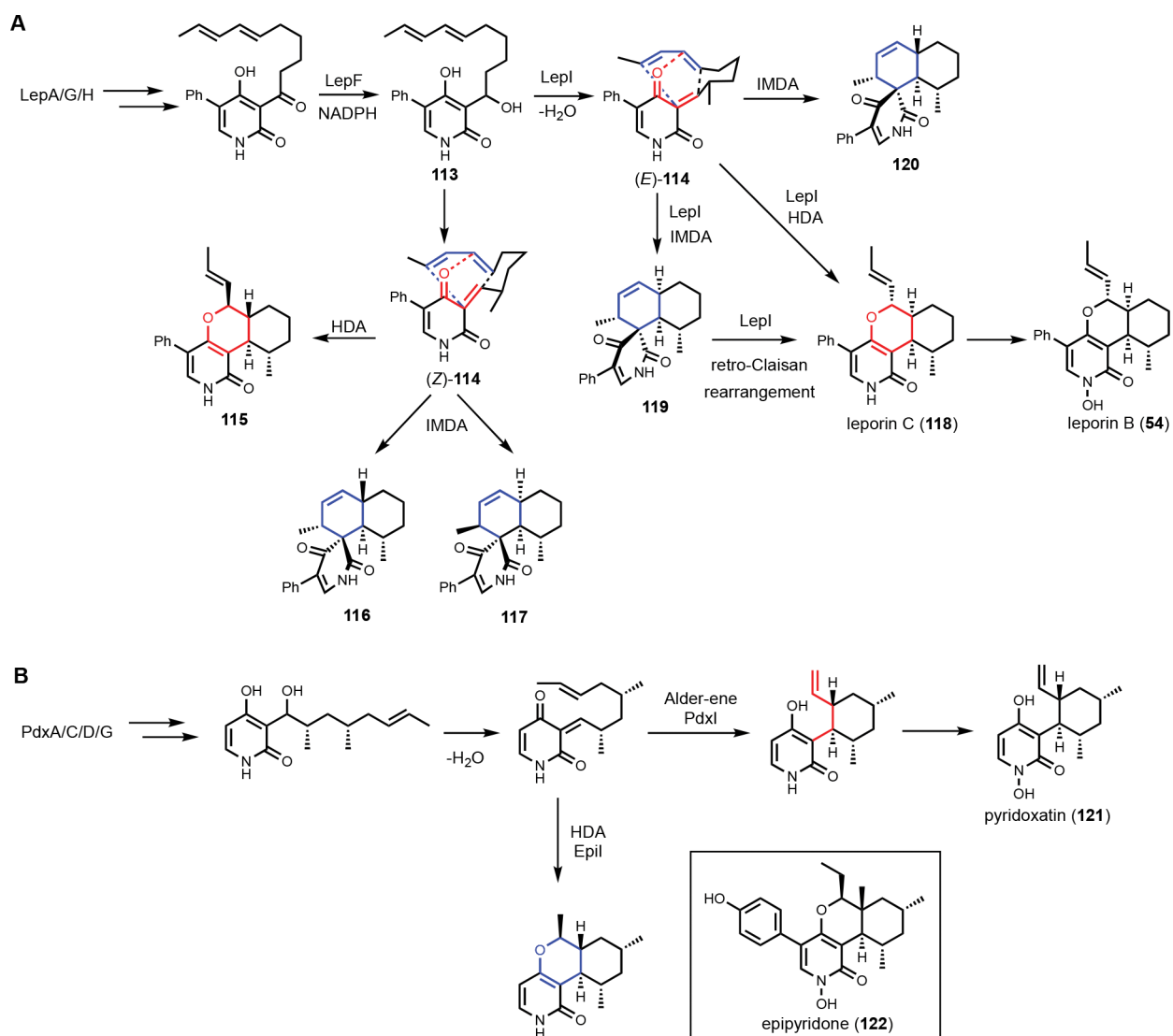
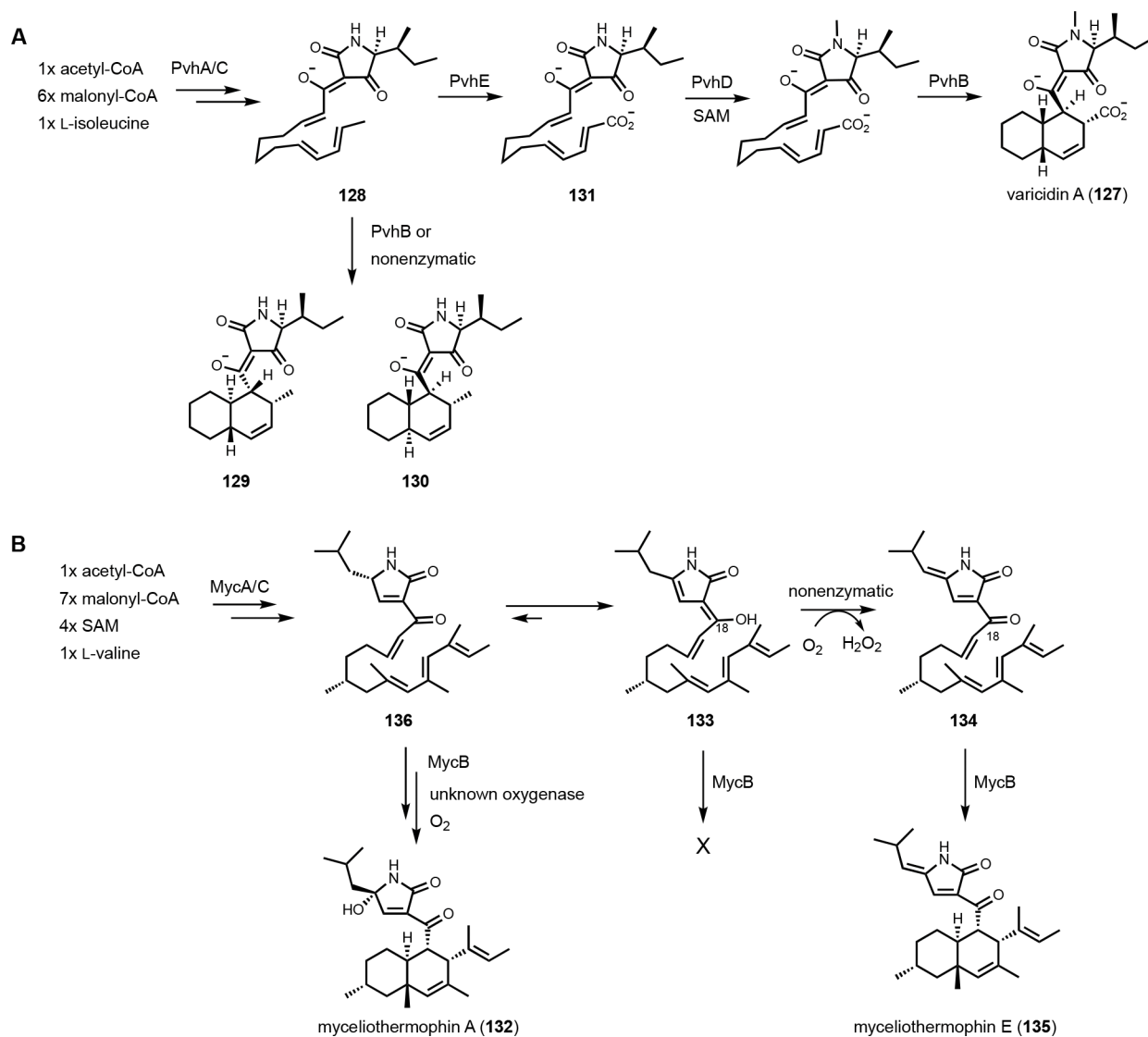
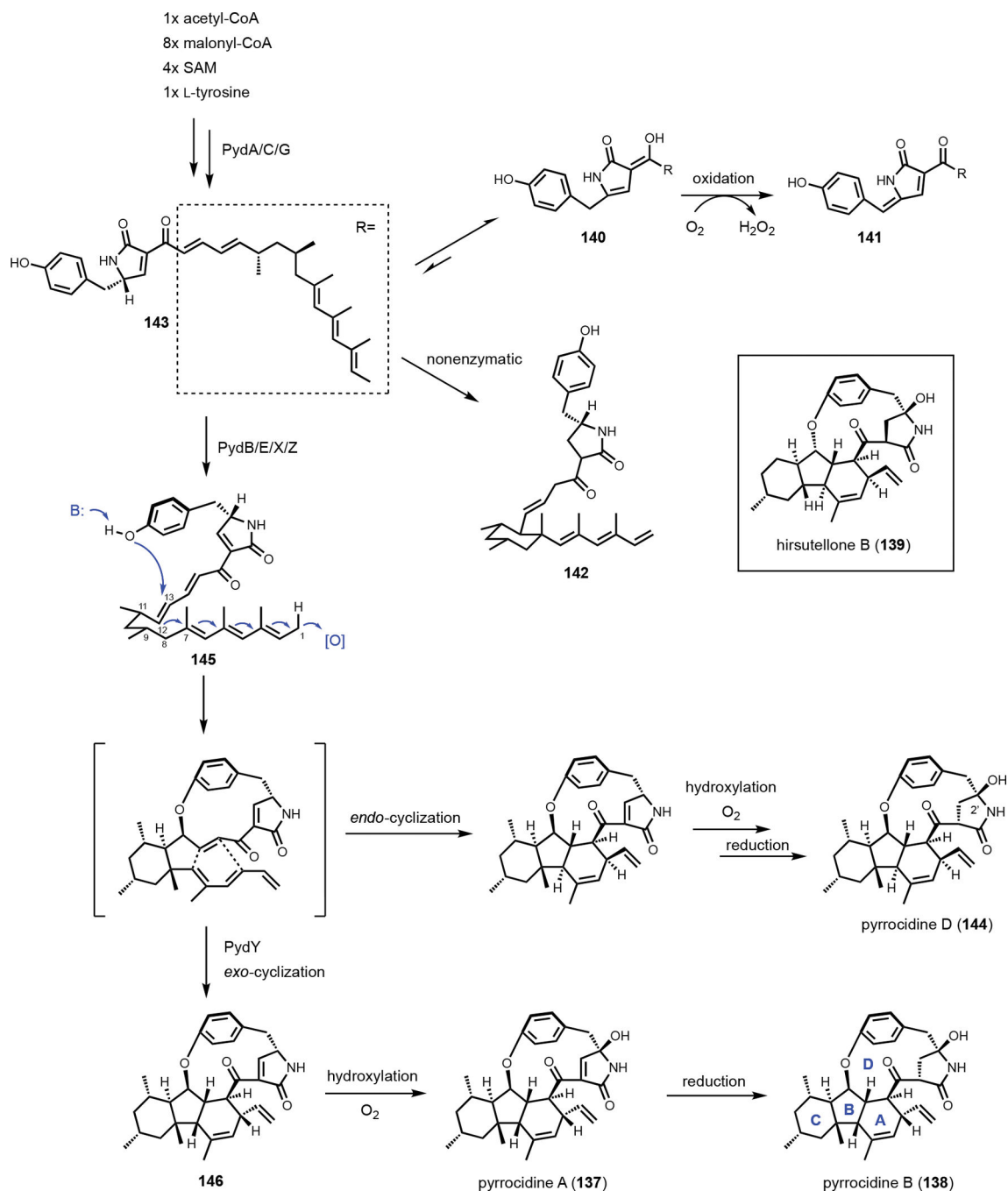


Figure 17. Pericyclases in 2-pyridones biosynthesis. **(A)** Biosynthetic pathway of leporin B. HDA reaction is highlighted in red, while IMDA reaction is labeled in blue. **(B)** two *O*-MT-like pericyclases, PdxI and EpiI, catalyze Alder-ene reaction and a different HDA reaction, respectively.

**Figure 18.**

Pericyclases in biosynthesis of decalin-containing natural products. **(A)** Biosynthetic pathway of varicidin. Carboxylation catalyzed by PvhE is required for the formation of *cis*-decalin product via a IEDDA mechanism. **(B)** Biosynthetic pathway of myceliothermophin. The shunt product **134** isolated from heterologous host allowed *in vitro* reaction to verify the function of MycB.

**Figure 19.**

Biosynthetic pathway of pyrrocidine. Expression of the putative Diels–Alderase PydB led to the formation of shunt product, pyrrocidine D, through *endo*-cyclization. PydY, which is predicted as hypothetical protein, is the pericyclase that catalyzes *exo*-selective cycloaddition to form the ring structure of pyrrocidine A and B.

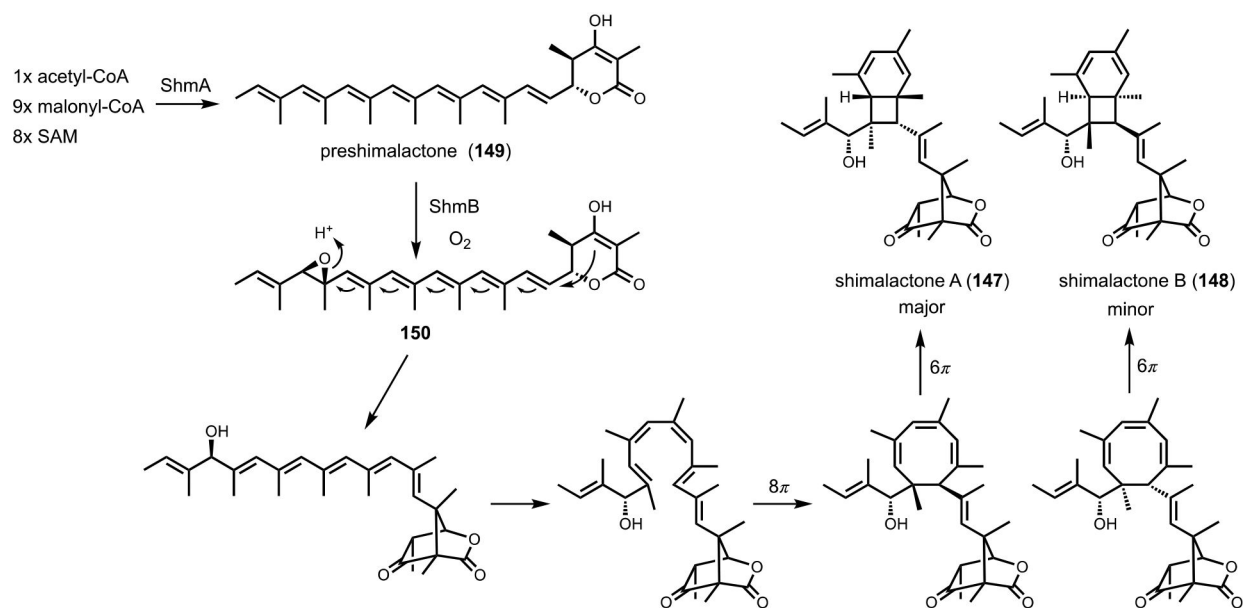
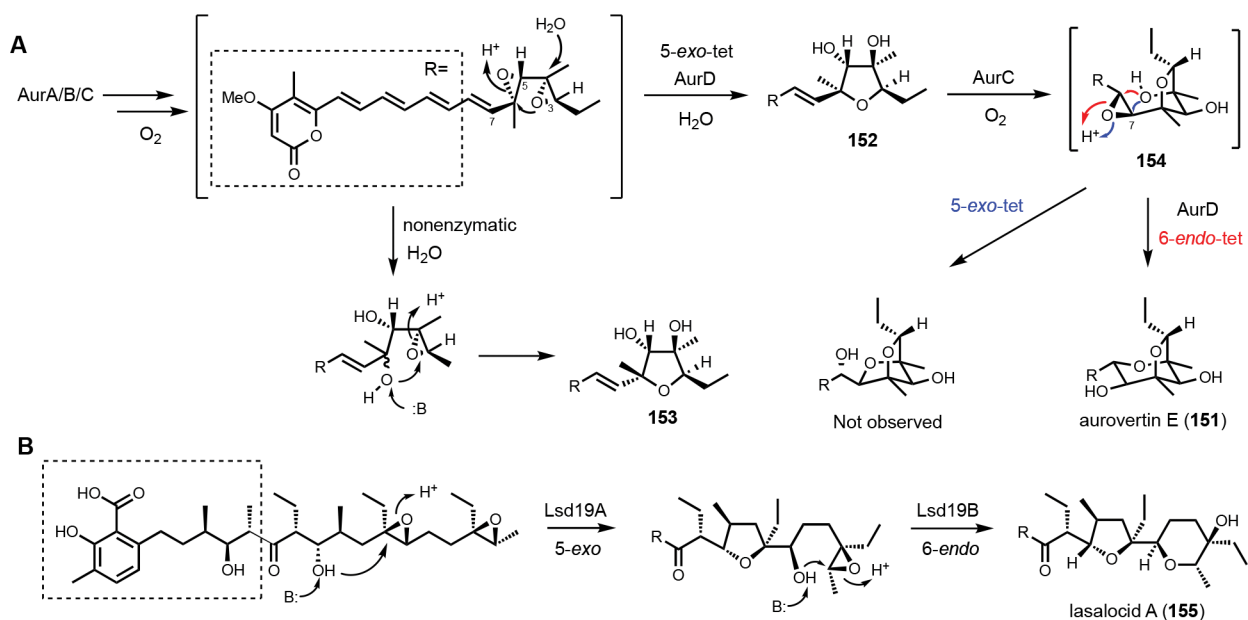


Figure 20. Biosynthetic pathway of Shimalactones involves nonenzymatic 8π - 6π electrocyclization to generate a diastereomeric pair of products.

**Figure 21.**

Epoxide opening reactions in polyether biosynthesis. **(A)** Sequential epoxide opening in biosynthesis of aurovertin. With the enzyme AurD, the final ring closure is via a 6-*endo*-tet reaction. **(B)** Sequential epoxide opening in biosynthesis of lasalocid A. The last 6-*endo*-tet reaction is classified as an anti-Baldwin's rule ring closure.

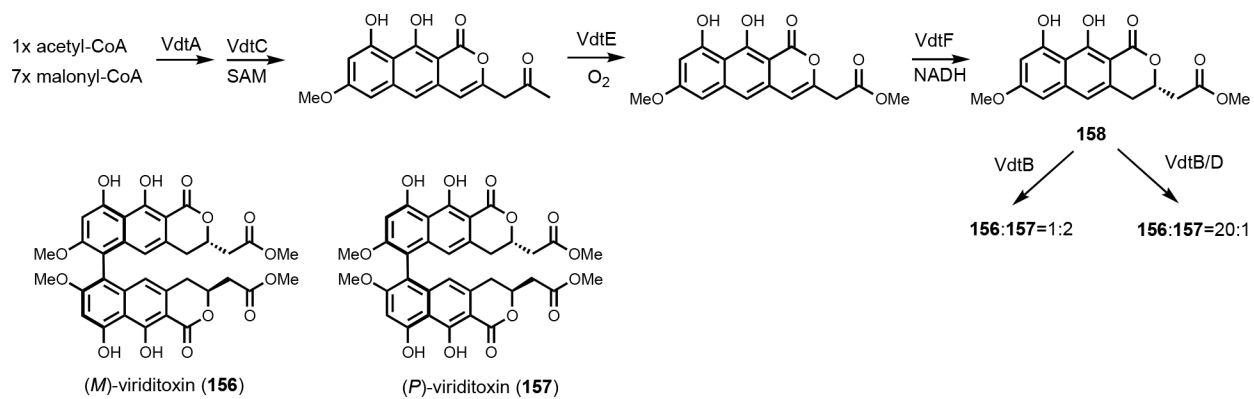
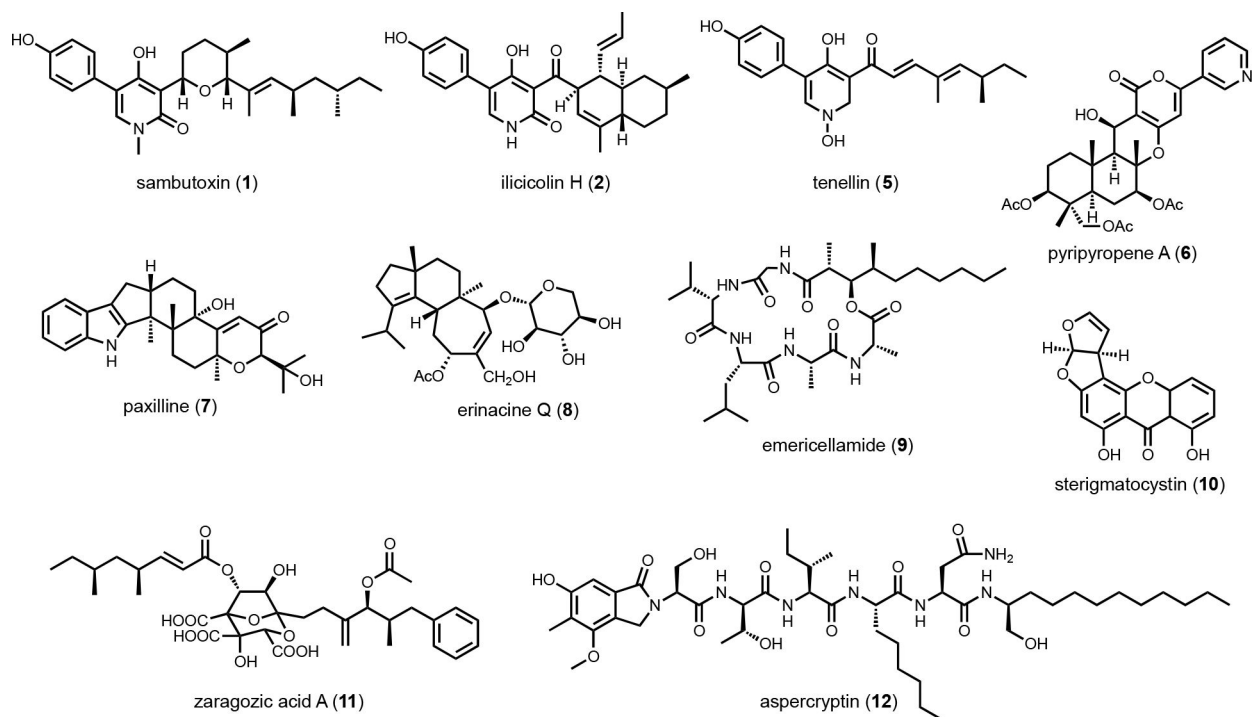
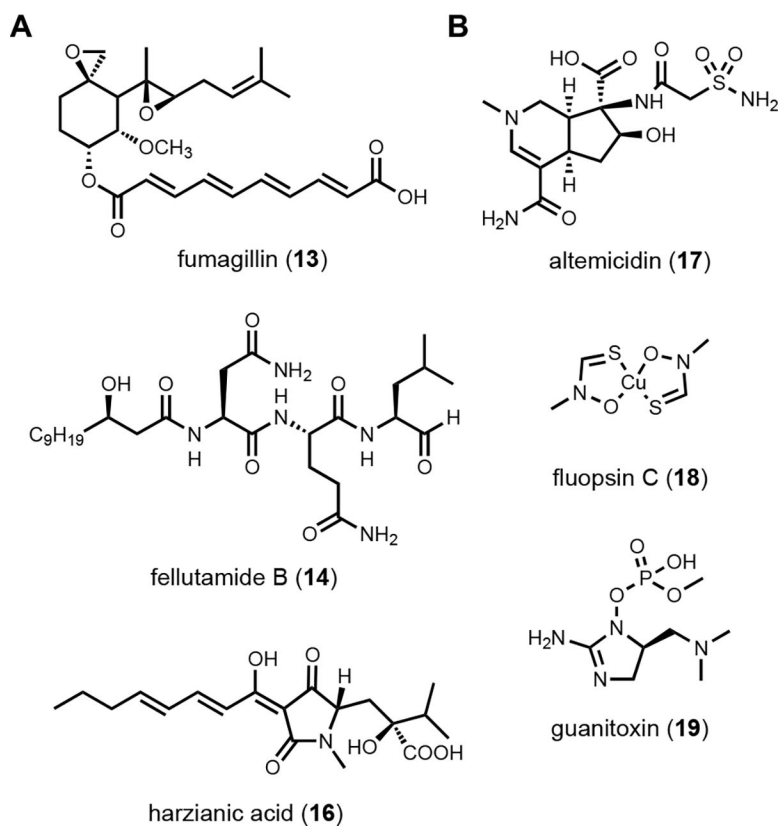


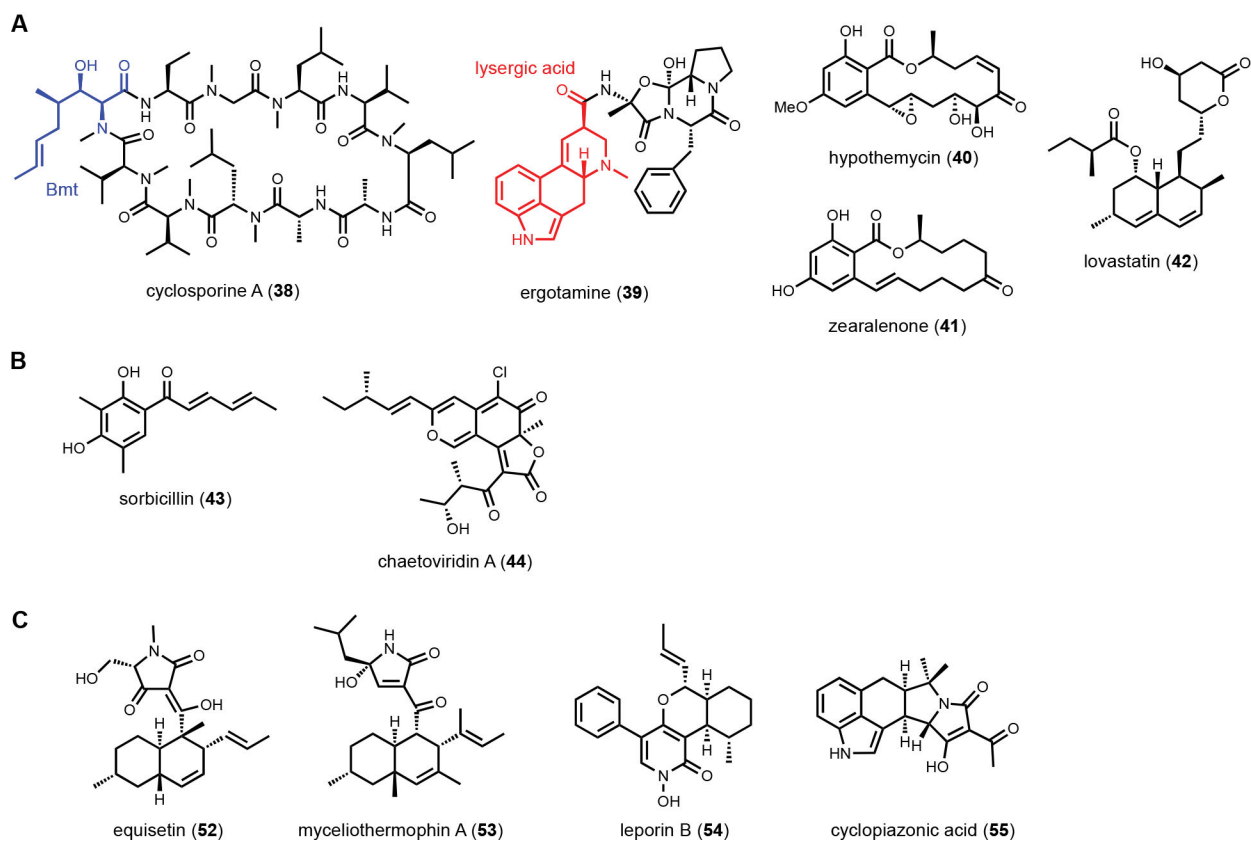
Figure 22.
Biosynthetic pathway of viriditoxin. Oxidative coupling reaction with the presence of VdtD led to high atroposelectivity for (*M*)-viriditoxin.



Scheme 1.
Metabolites discussed in section 2

**Scheme 2.**

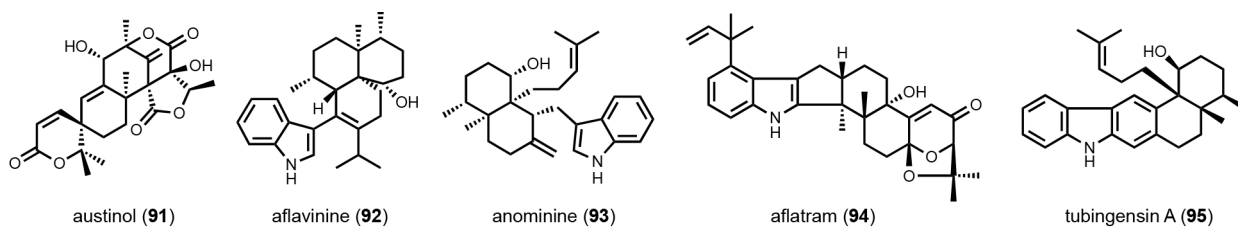
(**A**) Examples of metabolites found by SRE-guided genome mining. (**B**) Discoveries of natural product can be classified into unknown-known.



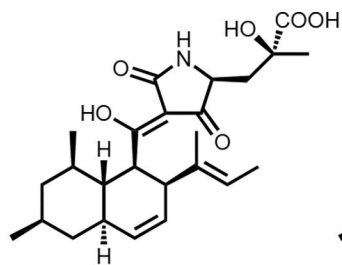
Scheme 3.

(A) Selected example of metabolites required collaboration of multiple core enzymes.

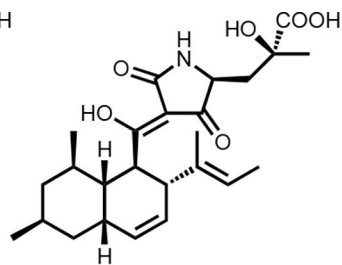
(B) Representative natural products synthesized by dual PKSs. (C) Selected PKS-NRPS metabolites using proteinogenic amino acid as building block.



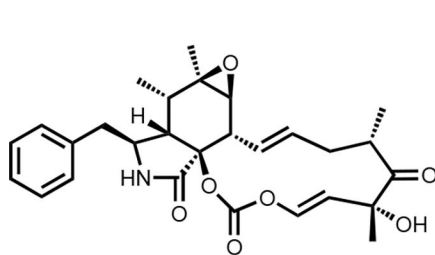
Scheme 4.
Examples of fungal indole diterpenes.



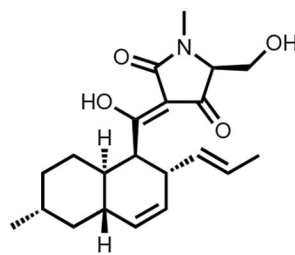
Sch210972 (123)



125



cytochalasin E (124)



equisetin (126)

Scheme 5.
Metabolites discussed in section 4.2.2