Organic & Biomolecular Chemistry

The Expanding Spectrum of Diketopiperazine Natural Product Biosynthetic Pathways Containing Cyclodipeptide Synthases

Journal:	*Organic & Biomolecular Chemistry*
Manuscript ID	OB-REV-12-2018-003063.R1
Article Type:	Review Article
Date Submitted by the Author:	20-Jan-2019
Complete List of Authors:	Borgman, Paul; Univ. of North Florida, Chemistry Lopez, Ryan; Univ. of North Florida, Chemistry Lane, Amy; Univ. of North Florida, Chemistry

Organic & Biomolecular Chemistry

REVIEW

The Expanding Spectrum of Diketopiperazine Natural Product Biosynthetic Pathways Containing Cyclodipeptide Synthases

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Paul Borgman,[a] Ryan D. Lopez [a] and Amy L. Lane* [a]

Microorganisms are remarkable chemists, with enzymes as their tools for executing multi-step syntheses to yield myriad natural products. Microbial synthetic aptitudes are illustrated by the structurally diverse 2,5-diketopiperazine (DKP) family of bioactive nonribosomal peptide natural products. Nonribosomal peptide synthetases (NRPSs) have long been recognized as catalysts for formation of DKP scaffolds from two amino acid substrates. Cyclodipeptide synthases (CDPSs) are more recently recognized catalysts of DKP assembly, employing two aminoacyl-tRNAs (aa-tRNAs) as substrates. CDPS-encoding genes are typically found in genomic neighbourhoods with genes encoding additional biosynthetic enzymes. These include oxidoreductases, cytochrome P450s, prenyltransferases, methyltransferases, and cyclases, which equip the DKP scaffold with groups that diversify chemical structures and confer biological activity. These tailoring enzymes have been characterized from nine CDPS-containing biosynthetic pathways to date, including four during the last year. In this review, we highlight these nine DKP pathways, emphasizing recently characterized tailoring reactions and connecting new developments to earlier findings. Featured pathways encompass a broad spectrum of chemistry, including the formation of challenging C-C and C-O bonds, regioselective methylation, a unique indole alkaloid DKP prenylation strategy, and unprecedented peptide-nucleobase bond formation. These CDPS-containing pathways also provide intriguing models of metabolic pathway evolution across related and divergent microorganisms, and open doors to synthetic biology approaches for generation of DKP combinatorial libraries. Further, bioinformatics analyses support that much unique genetically encoded DKP tailoring potential remains unexplored, suggesting opportunities for further expansion of Nature's biosynthetic spectrum. Together, recent studies of DKP pathways demonstrate the chemical ingenuity of microorganisms, highlight the wealth of unique enzymology provided by bacterial biosynthetic pathways, and suggest an abundance of untapped biosynthetic potential for future exploration.

Introduction

Secondary metabolites are produced by organisms from all domains of life, and encompass a diverse array of chemical structures that are evolutionarily crafted for biological function.[1-4] These natural products act as chemical communication signals between organisms,[5, 6] provide lead compounds for pharmaceutical and agrochemical development,[7, 8] act as chemical probes for deducing mechanisms of biological processes,[9] and inspire organic syntheses.[10]

Natural products with 2,5-diketopiperazine (DKP) scaffolds are structurally diverse and produced by a wide range of microorganisms (Fig. 1).[11, 12] The broad structural diversity of these cyclic dipeptides translates into an equally wide range of biological activities, including immunosuppressant, herbicidal, and anticancer effects exemplified by **1**-**6**.[12-17] The DKP ring confers increased structural rigidity and stability against proteolysis relative to acyclic peptide scaffolds, making DKPs attractive in pharmaceutical development.[12] Several reviews summarize the array of DKP chemical structures and biological activities.[11, 12, 18]

The diversity of DKP natural products arises from the enzyme-catalyzed incorporation of a wide variety of amino acids into the cyclic dipeptide scaffold as well as tailoring of this scaffold.[18-22] Enzymes from two families, the nonribosomal peptide synthetases (NRPSs) and cyclodipeptide synthases (CDPSs), are the major catalysts recognized for formation of the two peptide bonds in DKPs. These enzyme families are the focus of excellent recent reviews.[21, 23] NRPSs are long-studied, often massive (> 100-kDa) multimodular enzymes that act as molecular assembly lines, with each module typically catalyzing a cycle of amino acid adenylation, peptide bond formation, and optional tailoring reactions (e.g. methylation).[23, 24] DKPs originate from either the premature release of dipeptide intermediates from longer assembly lines or from NRPSs dedicated to cyclic dipeptide assembly.[19]

In contrast to NRPSs, CDPSs were first reported in 2009 and are small (~30-kDa) enzymes that utilize two aminoacyl-tRNAs (aa-tRNAs) as substrates for catalyzing formation of two peptide bonds between aminoacyl moieties to yield the DKP scaffold (Fig. 2a).[25] Aminoacyl-tRNAs are uncommon players in natural product assembly,[21, 26] making CDPSs intriguing members of Nature's biosynthetic repertoire. Structural and mechanistic studies support that the two aa-tRNA substrates bind at separate sites of CDPSs, with DKP formation occurring via a sequential ping-pong mechanism (Fig. 2b).[27-32]

Over 100 CDPSs have been functionally characterized, including more than 40 during the last two years.[25, 33-45] Most of these CDPSs exhibit some promiscuity with respect to at least one of their two aa-tRNA substrates, yielding multiple cyclic dipeptide products. Altogether, CDPSs have been demonstrated to catalyze assembly of

[a] Department of Chemistry, University of North Florida, 1 UNF Dr., Jacksonville, FL 32224, USA.
Email: Amy.Lane@unf.edu

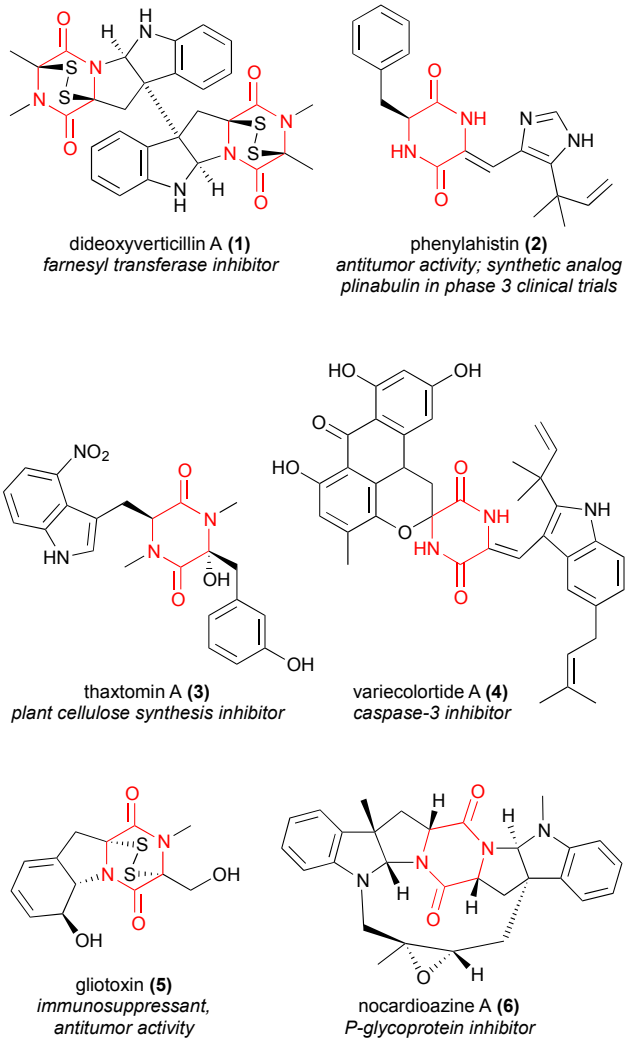

Fig. 1 Representative 2,5-diketopiperazine (DKP) natural products and bioactivities. The DKP core is shown in red. DKP natural products encompass diverse chemical structures and biological activities, as highlighted in italics.[12-17]

over 75 different cyclic dipeptides, spanning 18 of the 20 canonical L-amino acids. CDPSs were also recently demonstrated as catalysts for the formation of cyclic dipeptides containing approximately 25 different unnatural L-amino acids.[46]

Genes encoding multi-step biosyntheses of complex natural products are typically clustered on bacterial chromosomes, meaning that the discovery of one gene from a biosynthetic pathway typically facilitates discovery of others from the same pathway. Of the >100 functionally characterized CDPSs, nearly all are from bacterial gene clusters that also encode putative enzymes with hypothesized roles in DKP tailoring.[25, 33-45] Publicly available genome sequence databases contain over 700 additional putative CDPS-encoding genes, most of which also are clustered with predicted DKP tailoring genes.[47] In contrast, only nine biosynthetic pathways that include CDPSs have been experimentally characterized, yielding compounds **7**-**15** (Fig. 3).[29, 34, 36, 37, 41, 42, 44, 48-53] Four of these pathways were established within the last year, and yield some of the most structurally complex CDPS-derived natural products (i.e. **10-12**; **15**) known to date.[34, 41, 42, 44, 52, 53]

Herein, we highlight the nine functionally characterized biosynthetic pathways that employ CDPSs for cyclic dipeptide precursor assembly (Fig. 3). Our aim is to emphasize the most recently established pathways, along with providing overviews of earlier characterized pathways, and suggesting future directions of importance. We focus on enzymes that tailor CDPS-derived cyclic dipeptides rather than on CDPS-catalyzed reactions within these pathways, since other recent reviews provide a wealth of information on CDPS structures, functions, and mechanisms.[21, 26, 33] The recent surge in functional characterization of CDPSs and DKP pathways expands the spectrum of DKP biosynthesis, opening opportunities for application of DKP pathways in chemoenzymatic syntheses, synthetic biology, ecology, evolution, and more.

The spectrum of characterized biosynthetic pathways that contain CDPSs

Common themes among characterized pathways.

The nine characterized biosynthetic gene clusters that encode CDPSs are all from bacteria and range from under 2 kB to approximately 10 kB in size (Fig. 3),[29, 34, 36, 37, 41, 42, 44, 48-53] placing them among the smallest types of known natural product gene clusters. Their compactness arises from the small size of CDPSs relative to other common machinery (e.g. NRPSs[23] and polyketide synthases[54]) for natural product framework assembly and the relatively low number of tailoring enzymes encoded by most clusters. Tailoring enzymes

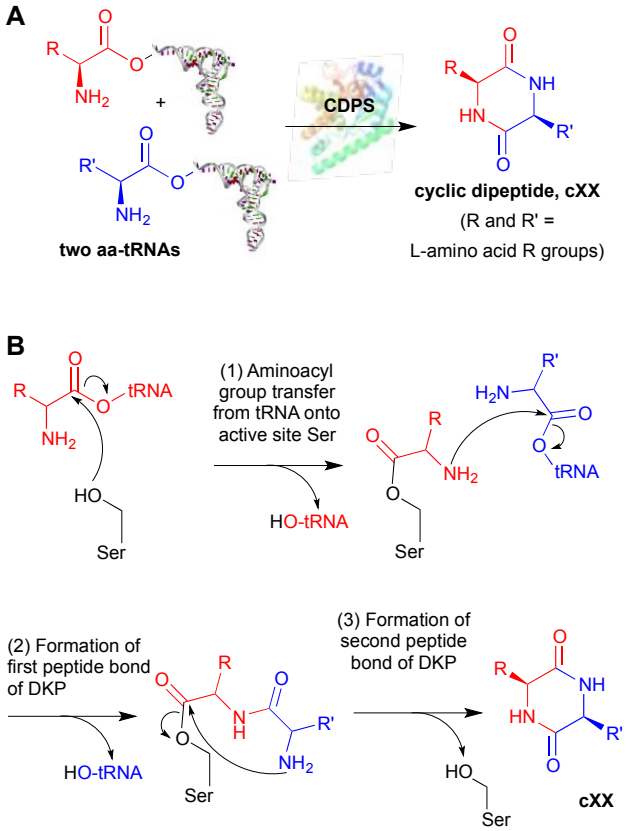

Fig. 2 CDPSs as catalysts for the biogenesis of cyclic dipeptides. (a) CDPSs utilize two aa-tRNAs as substrates for catalyzing cyclic dipeptide assembly. Cyclic dipeptide structures are commonly abbreviated as cXX, where X is the one letter abbreviation for an L-amino acid residue. (b) The CDPS catalytic mechanism entails initial covalent tethering of the aminoacyl moiety from the first aa-tRNA substrate (red) onto a conserved active site serine (Ser) residue. Nucleophilic attack of the amino nitrogen on the carbonyl carbon from the second aa-tRNA substrate (blue) yields the first peptide bond. The resulting enzyme-linked dipeptidyl intermediate then undergoes intramolecular peptide bond formation to yield the DKP group with concomitant release from the active site. The two aa-tRNA substrates bind at different sites of the CDPS.[27-32]

Organic & Biomolecular Chemistry

REVIEW

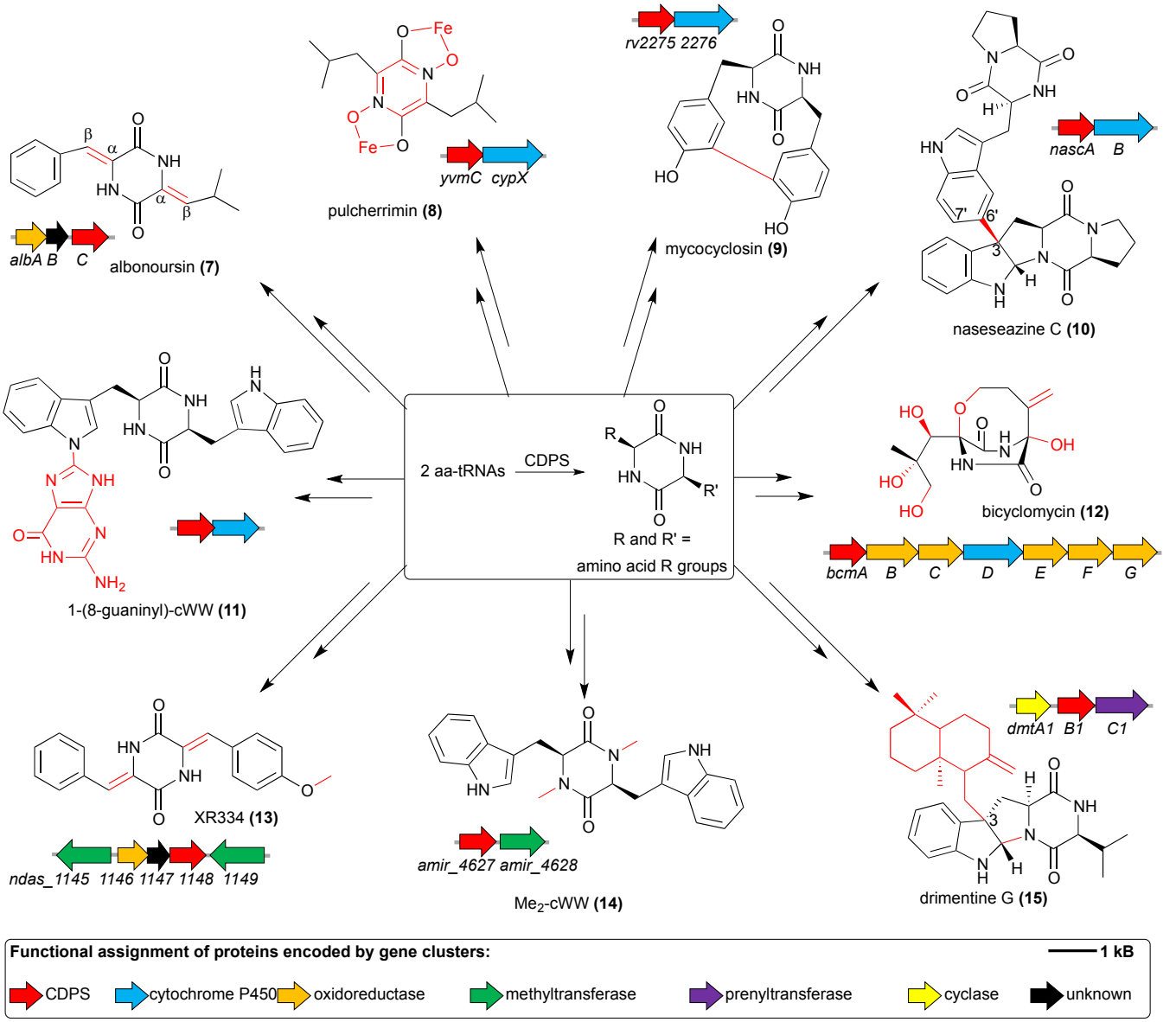

Fig. 3 Overview of representative DKPs produced by experimentally characterized biosynthetic pathways that contain CDPSs. At the beginning of each pathway, a CDPS catalyzes cyclic dipeptide formation from two aa-tRNA substrates (middle box). Tailoring reactions then transform cyclic dipeptide precursors into **7**-**15**; groups introduced via these tailoring steps are emphasized in red. A schematic of each biosynthetic gene cluster is provided alongside the chemical structure it encodes. These cluster schematics highlight experimentally characterized tailoring reactions yielding **7**-**15**.[29, 34, 36, 37, 41, 42, 44, 48-53]

catalyze reactions of both the cyclic dipeptide ring and amino acid R-groups, leading to much of the observed DKP structural diversity (Fig. 3).[29, 34, 36, 37, 41, 42, 44, 48-53] DKP diversity is further bolstered by the recurrent theme of promiscuity, in some cases beginning with CDPS-catalyzed formation of multiple cyclic dipeptide precursors and continuing with tailoring enzymes that accept multiple substrates to yield natural combinatorial libraries of DKPs. Additionally, homologous pathways from different bacteria often yield related DKPs, further exemplifying the combinatorial chemistry prowess of DKP pathways. Below, overviews of pathways yielding **7**-**15** highlight these biosynthetic themes.

Organic & Biomolecular Chemistry

REVIEW

The albonoursin pathway was the first-established CDPS-containing biosynthetic pathway and offers an example of pathway promiscuity and oxidative tailoring reactions.

The first recognized biosynthetic gene cluster featuring a CDPS was that encoding albonoursin (**7**, Fig. 3),[49] an antibacterial molecule isolated from *Streptomyces noursei* and related species.[56] Assembly of the cFL precursor of **7** is catalyzed by CDPS AlbC, which also yields a variety of other cyclic dipeptides as minor products.[25] A flavin-dependent oxidoreductase (AlbA) then catalyzes α,β-dehydrogenation at the phenylalanine hemisphere of cFL.[48] This enzyme subsequently catalyzes an equivalent reaction at the leucine hemisphere to yield di-α,β-dehydrogenated product **7**. AlbA exhibits promiscuity, catalyzing mono-α,β-dehydrogenation of a variety of cyclic dipeptides containing both aliphatic and aromatic side chains. The DKP ring appears an important determinant of substrate recognition, as dehydrogenation reactions were not observed for selected acylic dipeptides.[48] Intriguingly, the catalytic activity of AlbA requires AlbB, which lacks significant sequence similarity to characterized proteins and likely forms an active multimeric complex with AlbA.[49]

Pulcherrimin biosynthesis highlights oxidative tailoring reactions and different pathways yielding the same natural product.

Pulcherrimin (**8**) is a long-known and broadly distributed iron chelate, isolated from a variety of genera of yeast and *Bacillus* genus bacteria.[57, 58] In *Bacillus subtilis*, the CDPS YvmC catalyzes formation of major product cLL as well as minor cyclic dipeptide products containing one leucine residue.[25, 30] The *yvmC* gene is part of a two gene operon with *cypX* (Fig. 3),[59] which encodes a cytochrome P450 homologue. CypX is implicated as the catalyst for two N-hydroxylations and dehydrogenation of the cLL ring to yield aromatic pulcherriminic acid (**16**),[55] which is transformed into **8** via non-enzymatic chelation of iron(III) (Fig. 4).[57] Cryle and co-workers demonstrated that CypX catalyzes cLL ring hydroxylation *in vitro*, although the regiochemistry remains unknown. Their crystal structure of CypX revealed that it adopts the fold typical of P450s, featuring a predominantly α-helical structure and noncovalently linked heme moiety.[55]

With this evidence for the pulcherrimin pathway from *B. subtilis*, it appeared probable that yeasts assemble this DKP through a homologous pathway. Curiously, no homologues of *B. subtilis* pulcherrimin biosynthetic enzymes were revealed by bioinformatics analyses of sequenced yeast genomes. Krause et al. broadened the search by evaluating selected yeast genome sequences for putative proteins that contain conserved domains from cytochromes P450.[60] This led to identification of Pul2, a homologue of *Aspergillus* P450s that play roles in secondary metabolism, from the yeast *Kluyveromyces lactis*. Gene replacement mutations of genes from the *pul1-pul2* operon in *K. lactis* yeast, along with genetic complementation experiments, implicated Pul1 and Pul2 in biogenesis of **16** (Fig. 4).[60] Pul1 lacks sequence similarity to recognized enzyme families and is postulated to yield cLL as a precursor to **16**, although the product of Pul1 has yet to be directly confirmed. The pulcherrimin pathways from bacteria and yeast (Fig. 4) likely arose via convergent evolution, with catalysts from dissimilar evolutionary origins leading to the same natural products.[60]

The mycocyclosin pathway features a cytochrome P450 that catalyzes intramolecular DKP diaryl C-C linkage.

Intrigue about the two gene operon *rv2275-rv2276* (Fig. 3) from *Mycobacterium tuberculosis* was initiated largely by the discovery that Rv2276 is required for viability of this pathogen, sparking interest in this cytochrome P450 as an antibiotic drug target.[61] The

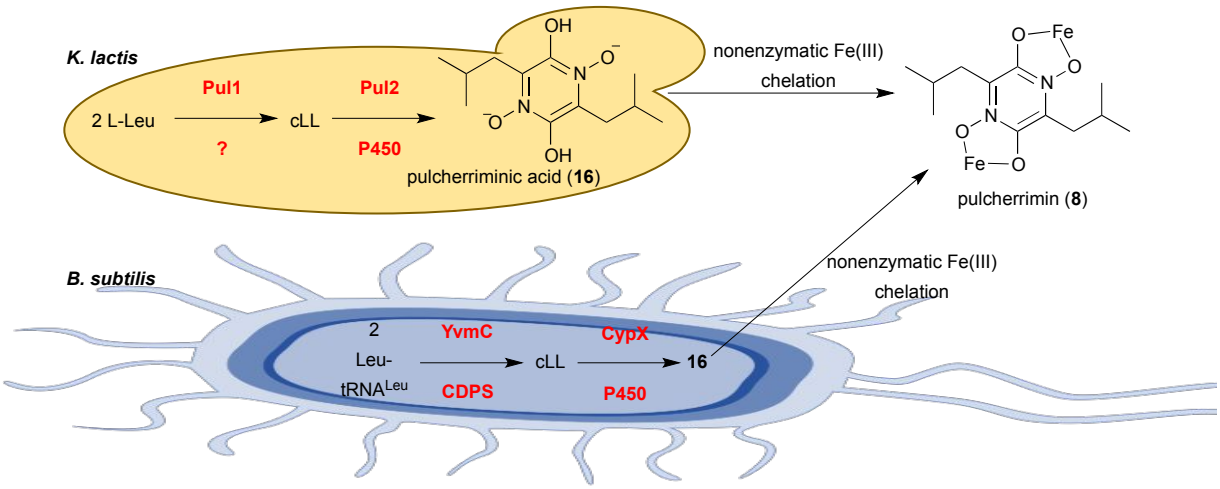

Fig. 4 Overview of pulcherrimin biosynthetic pathways from the yeast *K. lactis* and bacteria *B. subtilis*. In the pathway established from *B. subtilis*, formation of cLL is catalyzed by a CDPS followed by N-oxidation via a cytochrome P450 to yield **16**.[25, 30 59 55] This molecule is excreted into the environment, where it binds iron(III) to yield **8**. *K. lactis* also employs a two enzyme biosynthetic pathway, but neither of these enzymes exhibit sequence homology to the *B. subtilis* pathway.[60] This represents a likely example of convergent evolution of biosynthetic pathways from phylogenetically distant organisms.

Organic & Biomolecular Chemistry

REVIEW

function of P450 Rv2276 remained cryptic until Rv2275 was established as a CDPS that catalyzes formation of cYY as its major product, along with a handful of minor products containing tyrosine.[25, 29] This led to the hypothesis that P450 Rv2276 catalyzes functionalization of cYY. Indeed, *in vitro* assays revealed that Rv2276 catalyzes conversion of cYY to mycocyclosin (**9**).[50] The P450 Rv2276 appears substrate specific, as intramolecular diaryl C-C coupling was not observed for other aromatic cyclic dipeptide substrates.[62] X-ray crystal structures, molecular dynamics simulations, and spectroscopy studies supported that transformation of cYY into **9** proceeds through a bi-carbon radical mechanism.[50, 51, 63] These structural and functional characterization studies guided the development of synthetic inhibitors of Rv2276 as potential future tuberculosis therapies,[64] while the molecular basis for the essential physiological role of P450 Rv2276, cYY, and/or **9** remains under investigation.

A cytochrome P450 from the naseseazine pathway enabled development of biosynthetic tools for formation of challenging C-C bonds found in dimeric DKPs.

A large number of bioactive natural products, mostly from fungi, are homo- or heterodimeric DKPs (e.g. **1**).[13] This family of molecules poses formidable organic synthesis challenges, largely related to intermolecular C-C bond formation to link the two pyrroloindoline moieties of DKPs.[65] As a model for biosynthesis of dimeric pyrroloindoline DKPs, Tian and co-workers recently established the naseseazine (**10**) pathway as well as a facile *in vivo* system for formation of synthetically challenging C-C bonds of homo- and heterodimeric DKPs.[52]

Unlike fungal biosynthetic pathways that utilize NRPSs to form DKP precursors of dimeric DKPs, biogenesis of **10** from a marine-derived *Streptomyces* sp.[66] was linked to a CDPS-containing gene cluster (*nascA-nascB*, Fig. 3) via heterologous expression of this pathway in *S. albus*.[52] Sequence homology to characterized CDPSs implicated NascA in assembly of the cWP precursor of **10**, while functional characterization of purified recombinant NascB established it as the cytochrome P450 catalyst of intermolecular C-C bond formation between two cWP precursors. Like the mycocyclosin pathway P450 Rv2276 (Fig. 3), NascB was proposed to employ a biradical mechanism for C-C bond formation.

Discovery of the naseseazine pathway P450 opened doors for biosynthetic assembly of novel members of this dimeric natural product family. Tian and co-workers engineered *E. coli* as a cellular factory for production of pyrrolindoline DKP homo- and heterodimers.[52] Supplementation of *E. coli* cultures expressing NascB with individual or combinations of synthetic indole alkaloid cyclic dipeptides resulted in isolation of approximately 30 novel DKP homo- and heterodimers, corresponding to intermolecular bond formation between indole-C6' or -C7' and indole-C3 of the two DKP monomers (e.g. **10**). DKP dimer products featured a variety of natural and unnatural amino acids, including some D-amino acids, highlighting the promiscuity of NascB in DKP substrate recognition. These results highlight the promise of DKP pathway enzymes as tools for the generation of structurally unique molecules that are challenging to access via organic syntheses.

A DKP pathway provided the first example of peptide functionalization with a nucleobase, extending the functional spectrum of P450 tailoring enzymes.

Electron-rich indole moieties are common targets of a variety of biosynthetic tailoring reactions, and the resulting functionalized indole groups are valued for their fascinating structures and essential contributions to the biological activities of many natural products (e.g. **1**; **3**; **4**; **6**).[67] The mining of *Streptomyces* genomes for novel CDPS-containing biosynthetic pathways recently resulted in discovery of an unprecedented indole tailoring reaction, in which a P450 catalyzes linkage of a DKP indole nitrogen with a guanine carbon.[41]

Bioinformatics analyses supported homologous putative operons encoding a CDPS and P450 (Fig. 3) from three strains of *Streptomyces*. Expression of each CDPS in *E. coli* resulted in accumulation of cWW as the sole detectable DKP product.[33, 35] Introduction of the two gene CDPS-P450 operon into *S. coelicolor* resulted in production of a molecule identified as 1-(8-guaninyl)-cWW (**11**) by 1D and 2D NMR analyses.[41] *In vitro* assays further confirmed this finding, demonstrating P450-catalyzed formation of the C-N linkage between cWW and guanine substrates to yield **11**. This P450 shares sequence similarity with both NascB and Rv2276, which catalyze C-C bond formation, and likely employs a similar radical mechanism. Future structural studies may further illuminate this mechanism and establish determinants of DKP and nucleobase substrate recognition. By providing the first example of nucleobase tailoring of a peptide natural product from any biosynthetic origin (i.e. RiPP, NRPS, or CDPS), this work expands Nature's biosynthetic toolkit and highlights the promise of CDPS-containing pathways as sources of novel biosynthetic transformations.

The bicyclomycin pathway is the most complex CDPS-containing biosynthetic pathway characterized to date and a model for the study of pathway evolution.

Bicyclomycin (**12**) is a structurally compact and remarkably oxidized DKP.[68] It is the only known selective inhibitor of transcription termination factor *rho*, necessary for survival of many bacteria.[69] Understanding of the biosynthetic pathway to **12** was limited primarily to precursor feeding studies until 2018,[70] when the biogenesis mystery was solved by three groups in parallel. Genome sequencing of bicyclomycin-producing actinomycete strains by each group revealed a ~9 kB locus (*bcm*) predicted to encode a CDPS, five α-ketoglutarate-dependent dioxygenases, and a cytochrome P450 monooxygenase (Fig. 3).[34, 42, 44]

Introduction of the *bcm* cluster into heterologous host *S. coelicolor* by Vior and co-workers enabled production of **12**, supporting the role of some or all genes from this cluster in

Organic & Biomolecular Chemistry

REVIEW

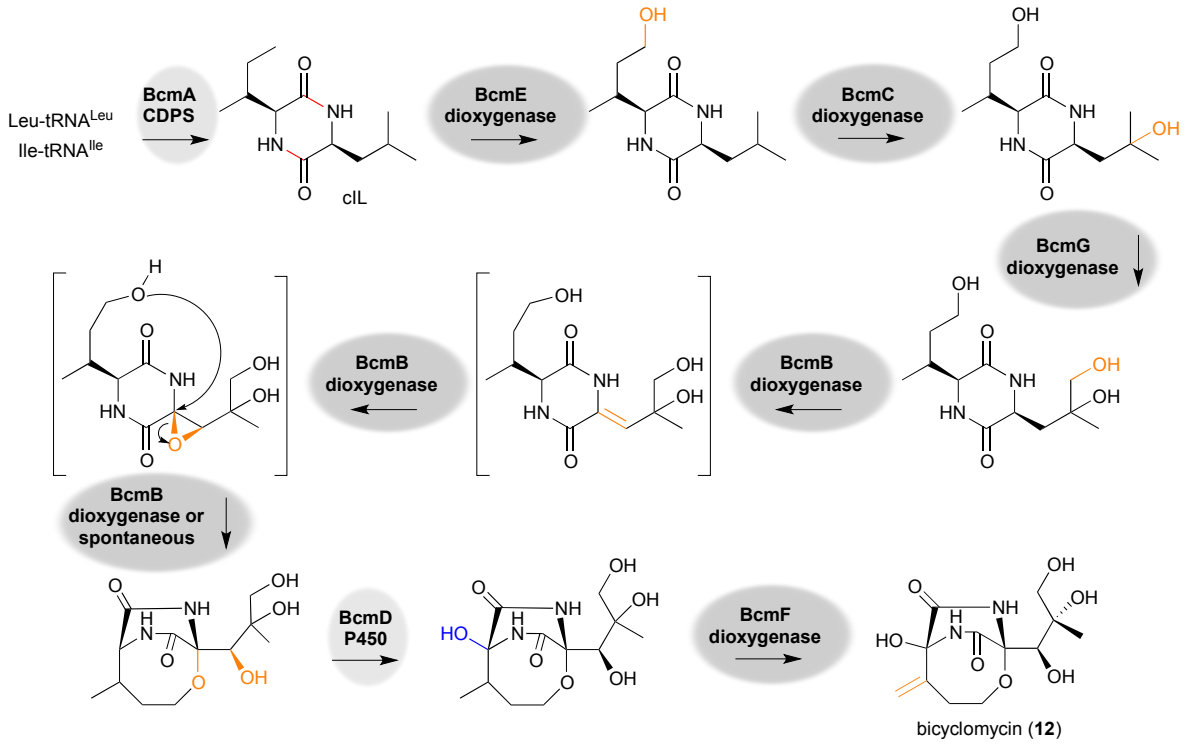

Fig. 5 Proposed bicyclomycin biosynthetic pathway, featuring a cascade of oxidative transformations. This biosynthetic route was proposed by Meng and co-workers based on *in vitro* reconstitution of the entire pathway.[34] Encompassing a series of six enzymes catalyzing oxidation reactions, this represents the longest CDPS-containing pathway characterized to date.

biosynthesis of **12**.[44] Concurrently, both Patteson et al. and Meng et al. undertook the challenge of establishing the pathway to **12** by functional characterization of recombinant enzymes from the *bcm* cluster (Fig. 3).[34, 42] BcmA was established as a CDPS, catalyzing formation of the cIL scaffold. The pathway then proceeds through hydroxylation of three unactivated sp^3 carbons catalyzed by BcmE, BcmC, and BcmG. This is followed by α,β-dehydrogenation and epoxidation at the leucine hemisphere catalyzed by BcmB, with potentially spontaneous nucleophilic attack from the isoleucine terminal hydroxy group on the epoxide to yield an 8-membered ring. Finally, hydroxylation of a DKP ring carbon by P450 BcmD and exo-methylene formation via BcmF-catalyzed dehydrogenation complete the pathway to **12** (Fig. 5).[34, 42] With seven enzymes playing confirmed roles in DKP scaffold assembly and oxidative tailoring reactions, the route to **12** is the most elaborate CDPS-dependent pathway characterized to date.

The *bcm* gene cluster also highlights DKPs as exemplars for secondary metabolite pathway evolution. Observation of intact gene clusters across distantly related phyla is uncommon, yet bioinformatics analyses revealed largely syntenic *bcm* gene cluster homologues from the genome sequences of over 700 Gram-positive Actinobacteria and Gram-negative Proteobacteria.[44] The majority of these clusters were found within *Pseudomonas* genus Proteobacteria, including pathogenic *Pseudomonas aeruginosa* strains. Expression of *P. aeruginosa bcm* gene clusters in heterologous hosts resulted in production of **12**,[42, 44] suggesting that these pathways are functional. This finding raises additional questions about whether **12** is produced in natural habitats, and the possibility of this antibiotic as a mediator of biotic interactions within the human microbiome during *P. aeruginosa* infections.

Phylogenetic analyses revealed that *bcm* clusters from Gram-negative and Gram-positive bacteria group into different clades, with greater sequence divergence within Gram-positive clades.[44] This suggests that the *bcm* gene cluster likely originated in Gram-positive bacteria and was more recently introduced into Gram-negative bacteria via horizontal gene transfer. Additionally, the *bcm* cluster is typically flanked by predicted integrases, transposes, and other genetic elements responsible for gene cluster mobilization, providing additional circumstantial evidence for horizontal gene transfer as the origin of the broad distribution of the *bcm* cluster.[44]

Methyltransferases tailor both the DKP ring and amino acid R-groups.

S-adenosylmethionine (SAM) dependent methyltransferases are common tailoring enzymes across the natural product spectrum.[71]

Classic methyltransferases catalyze S_N2-type nucleophilic substitution reactions, in which a nucleophilic atom from the natural product substrate attacks the electrophilic methyl carbon of SAM as its methyl-sulfur bond breaks. Methyltransferase homologues are commonly encoded within putative CDPS gene clusters,[47] yet methyltransferases from only two of these clusters have been characterized to date. One leads to methylated members of the nocazine/XR334 (e.g. **13**) family and the other catalyzes DKP N-methylation of cWW to yield **14**.[36, 37]

Nocazine/XR334 family members have been isolated from multiple actinomycetes and include at least ten different molecules from cYY, cFF, and cFY precursors. Representatives of this natural combinatorial library feature N- and/or O-methylation of the DKP ring, O-methylation of tyrosine, and α,β-dehydrogenation.[72, 73] As a model for the pathway leading to this family, Giessen and co-workers characterized enzymes of the nocazine/XR334 gene cluster from *Nocardiopsis dassonvillei* (Fig. 3).[36] Assembly of the cyclic dipeptide scaffold is catalyzed by CDPS Ndas_1148. Ndas_1146 and Ndas_1147 are homologous to the oxidoreductase AlbA and hypothetical protein AlbB, respectively, from the albonoursin pathway (Fig. 3) and also catalyze α,β-dehydrogenation. Incubation of recombinant Ndas_1149 with a variety of cyclic dipeptides and SAM revealed that this methyltransferase exclusively catalyzes O-methylation of DKP tyrosine residues. The other methyltransferase homologue from the pathway, Ndas_1145, evaded experimental characterization but likely catalyzes DKP N- and O-methylation observed in some nocazine/XR334 family members.

Characterization of an orphan biosynthetic gene cluster from *Actinosynnema mirum* led to identification of an N-methylation catalyst of DKP rings.[37] Bioinformatics analyses of the *amir_4627 amir_4628* gene cluster suggested that it encoded a and methyltransferase (Fig. 3). Functional characterization of Amir_4627 revealed that it exclusively catalyzes cWW formation. Co-expression of this CDPS and methyltransferase candidate Amir_4628 as well as *in vitro* assays with purified Amir_4628 resulted in production of cWW derivatives with methylation of one or both DKP nitrogens (e.g. **14**), establishing Amir_4628 as an N-methyltransferase. The promiscuity of Amir_4628 was evaluated by co-expression in *E. coli* with AlbC, the albonoursin (**7**) pathway CDPS that catalyzes formation of cFX (X = variable amino acid). Evaluation of N-methyl DKP products from these cultures revealed a variety of cFX's as suitable substrates for Amir_4628. This *in vivo* biosynthetic pathway mixing-and-matching strategy not only provided a creative approach for evaluating methyltransferase promiscuity, but also suggests the promise of analogous biosynthetic approaches for further expanding the spectrum of DKP chemical diversity.

An unusual prenyltransferase from the drimentine pathway expands Nature's indole alkaloid prenylation strategies.

Indole prenyltransferases are a large family of homologous tailoring enzymes recognized for normal and reverse prenylation of a variety of indole alkaloid substrates, including DKPs that contain tryptophan.[74] Surprisingly, recent characterization of the drimentine biosynthetic pathway revealed a catalyst of indole alkaloid DKP prenylation that lacks significant homology to this enzyme family and instead is most homologous to phytoene synthases.[53]

Yao and co-workers observed that genome sequences from three *Streptomyces* strains shared a homologous locus predicted to encode a CDPS (DmtB), membrane-associated terpene cyclase (DmtA), and phytoene synthase (DmtC, Fig. 3.).[53] Introduction of partial gene cluster *dmtB1-dmtC1* from one of these strains into *S. coelicolor* resulted in production of cWP, cWV, and cWL derivatives with a farnesyl group at indole C-3, dubbed pre-drimentines (e.g. **17**, Fig. 6). This supported DmtB as a CDPS and DmtC, suggested by bioinformatics as a phytoene synthase, as a prenyltransferase. *In vitro* enzyme assays confirmed these functions of DmtB and DmtC and demonstrated that each exhibits promiscuity. Expression of the complete *dmtA1-C1* locus in *S. coelicolor* resulted in production of drimentine G (**15**), corresponding to cyclization of the farnesyl group of **17** (Fig. 6) and supporting DmtA as a farnesyl cyclase.

Intriguingly, some drimentine pathways may offer additional biosynthetic tricks beyond those encoded by *dmtA-dmtC*.[53] Gene clusters from two of the three *Streptomyces* strains harboring *dmtA-dmtC* additionally encode putative oxidases homologous to those catalyzing α,β-dehydrogenation in the albonoursin and nocazine/XR334 pathways as well as methyltransferase homologs. This suggests that further genome mining may unveil additional biosynthetic transformations to expand the drimentine family of natural products. The similarities and differences between *dmt* gene clusters from different strains also suggests these clusters as an interesting model for exploration of biosynthetic pathway evolution.

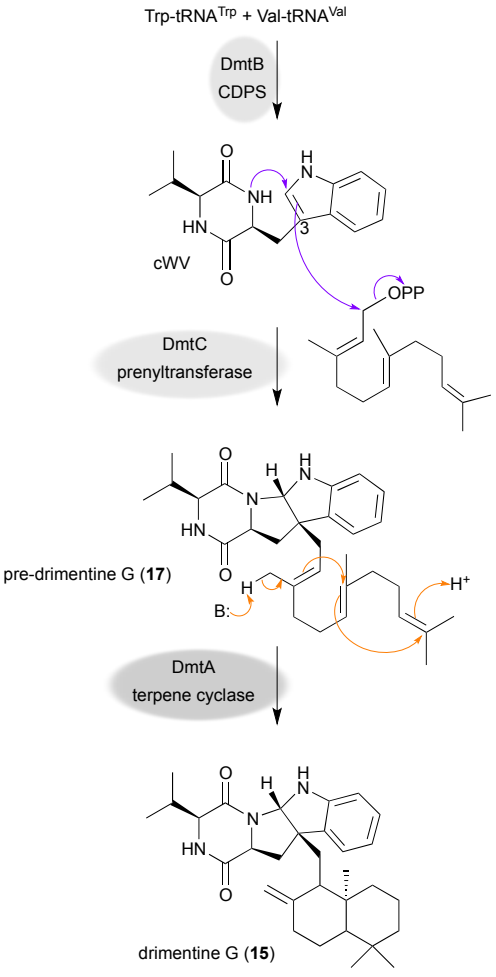

Fig. 6 Drimentine pathway, including unique prenylation biosynthetic machinery. In the pathway proposed by Yao et al.,[53] formation of the cWV precursor of **15** is catalyzed by CDPS DmtB. Indole C-3 farnesylation is catalyzed by the prenyltransferase DmtC, which lacks sequence homology to the large and well-studied family of indole prenyltransferases. Finally, cyclization of the prenyl group of **17** is catalyzed by DmtA to yield **15**.

Conclusions and Future Directions

The aptitude of enzymes as catalysts for assembly of elaborate natural products amazes and inspires, as illustrated by the above vignettes for nine characterized biosynthetic pathways that contain CDPSs. These pathways encompass a wide breadth of DKP tailoring reactions, including formation of challenging C-C (e.g. **9**-**10**) and C-O bonds (e.g. **12**), regioselective methylation (e.g. **13**-**14**), a unique prenylation strategy (e.g. **15**), and unprecedented bonds between groups such as peptides and nucleobases (e.g. **11**).[29, 34, 36, 37, 41, 42, 44, 48-52]

While these studies significantly expand the spectrum of DKP biosynthetic pathways, bioinformatics analyses suggest that much of the chemistry possible from CDPS-containing pathways remains untapped. Skinnider and co-workers recently developed an automated bioinformatics pipeline to predict products of CDPSs and tailoring enzymes from these pathways, and applied this tool to evaluate ~700 putative CDPS-containing biosynthetic pathways.[47] CDPSs from these pathways were predicted to catalyze formation of at least 25 cyclic dipeptides in addition to the approximately 75 previously reported. These analyses also supported that a large number of tailoring enzyme types remain unexplored, including ones predicted to catalyze cyclic dipeptide acetylation, sulfonation, and glycosylation. A variety of predicted cytochromes P450 and other oxidase families were also noted in pathways and are phylogenetically distinct from previously characterized DKP tailoring enzymes. Considering the assortment of unique transformations represented within the nine established CDPS-containing pathways (Fig. 3),[29, 34, 36, 37, 41, 42, 44, 48-53] characterization of these cryptic DKP pathways holds significant potential for expanding the breadth of Nature's chemistry toolkit.

The establishment of CDPSs and associated tailoring enzymes as catalysts for assembly of functionalized DKPs (Fig. 3) opens doors for synthetic biology approaches to DKP assembly. This synthetic biology promise is suggested by the small size of CDPSs relative to NRPS catalysts of cyclic dipeptide formation,[25] CDPS-catalyzed production of a wide breadth of cyclic dipeptides including ones with unnatural amino acids,[33, 46] and the relatively high *in vivo* yields of cyclic dipeptides (>100-200 mg/L) observed by heterologous expression of some CDPSs in *E. coli.*[25, 33, 35, 38, 40, 43] Further, several tailoring enzymes featured in this review, including P450s,[52] prenyltransferases,[53] and methyltransferases,[36, 37] exhibited promiscuity with respect to DKP substrate. The mixing-and-matching of CDPSs with promiscuous tailoring enzymes from multiple CDPS- and/or NRPS-containing pathways offers particular promise for combinatorial biosynthetic chemistry approaches to yield libraries of novel DKPs. Together, the untapped potential of CDPS pathways as sources of new DKPs along with the synthetic biology promise of CDPS pathways suggests a bright future for Nature's arsenal of bioactive DKP chemical diversity.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Research in the Lane group is supported by NSF RUI grant #1709655, as well as University of North Florida (UNF) Transformational Learning Opportunity Awards and Academic Affairs Scholarship grants. PB acknowledges support from a UNF Office of Undergraduate Research SURF grant, and RL a UNF Presidential Undergraduate Research Fellowship.

Notes and references

1. A. Schueffler and T. Anke, *Nat Prod Rep*, 2014, **31**, 1425-1448.
2. J. Clardy, M. A. Fischbach and C. T. Walsh, *Nat Biotechnol* 2006, **24**, 1541-1550.
3. J. W. Blunt, A. R. Carroll, B. R. Copp, R. A. Davis, R. A. Keyzers and M. R. Prinsep, *Nat Prod Rep*, 2018, **35**, 8-53.
4. C. T. Walsh and M. A. Fischbach, *J. Am. Chem. Soc.*, 2010, **132**, 2469-2493.
5. P. Spiteller, *Nat Prod Rep*, 2015, **32**, 971-993.
6. M. F. Traxler and R. Kolter, *Nat Prod Rep*, 2015, **32**, 956-970.
7. D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2016, **79**, 629-661.
8. J. Meinwald, *J. Nat. Prod.*, 2011, **74**, 305-309.
9. E. E. Carlson, *ACS Chem Biol*, 2010, **5**, 639-653.
10. K. C. Nicolaou, *J. Org. Chem.*, 2005, **70**, 7007-7027.
11. M. P. de Carvalho and W. R. Abraham, *Curr Med Chem*, 2012, **19**, 3564-3577.
12. A. D. Borthwick, *Chem. Rev.*, 2012, **112**, 3641-3716.
13. B. W. Son, P. R. Jensen, C. A. Kauffman and W. Fenical, *Nat. Prod. Lett.*, 1999, **13**, 213-222.
14. W. L. Wang, T. J. Zhu, H. W. Tao, Z. Y. Lu, Y. C. Fang, Q. Q. Gu and W. M. Zhu, *Chem. Biodivers.*, 2007, **4**, 2913-2919.
15. R. Raju, A. M. Piggott, X.-C. Huang and R. J. Capon, *Org Lett*, 2011, **13**, 2770-2773.
16. R. R. King, C. H. Lawrence and L. A. Calhoun, *J. Agri. Food Chem.* , 1992, **40**, 834-837.
17. K. Kanoh, S. Kohno, T. Asari, T. Harada, J. Katada, M. Muramatsu, H. Kawashima, H. Sekiya and I. Uno, *Bioorg Med Chem Lett*, 1997, **7**, 2847-2852.
18. A. Mishra, J. Choi, S.-J. Choi and K.-H. Baek, *Molecules*, 2017, **22**, 1796.
19. B. Gu, S. He, X. Yan and L. Zhang, *Appl Microbiol Biotechnol*, 2013, **97**, 8439-8453.
20. T. Giessen and M. Marahiel, *Int. J. Mol. Sci.*, 2014, **15**, 14610-14631.
21. M. Moutiez, P. Belin and M. Gondry, *Chem Rev*, 2017, **117**, 5578-5618.
22. T. W. Giessen and M. A. Marahiel, *FEBS Lett.*, 2012, **586**, 2065-2075.
23. J. A. Payne, M. Schoppet, M. H. Hansen and M. J. Cryle, *Mol Biosyst*, 2016, **13**, 9-22.
24. K. J. Labby, S. G. Watsula and S. Garneau-Tsodikova, *Nat Prod Rep*, 2015, **32**, 641-653.
25. M. Gondry, L. Sauguet, P. Belin, R. Thai, R. Amouroux, C. Tellier, K. Tuphile, M. Jacquet, S. Braud, M. Courçon, C. Masson, S. Dubois, S. Lautru, A. Lecoq, S.-i. Hashimoto, R. Genet and J.-L. Pernodet, *Nat. Chem. Biol.*, 2009, **5**, 414-420.
26. E. C. Ulrich and W. A. van der Donk, *Curr Opin Chem Biol*, 2016, **35**, 29-36.
27. M. Moutiez, E. Schmitt, J. Seguin, R. Thai, E. Favry, P. Belin, Y. Mechulam and M. Gondry, *Nat Comm*, 2014, **5**, 5141.
28. L. Sauguet, M. Moutiez, Y. Li, P. Belin, J. Seguin, M. H. Le Du, R. Thai, C. Masson, M. Fonvielle, J. L. Pernodet, J. B. Charbonnier and M. Gondry, *Nucleic Acids Res.*, 2011, **39**, 4475-4489.

29. M. W. Vetting, S. S. Hegde and J. S. Blanchard, *Nat. Chem. Biol.*, 2010, **6**, 797-799.
30. L. Bonnefond, T. Arai, Y. Sakaguchi, T. Suzuki, R. Ishitani and O. Nureki, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 3912-3917.
31. M. Moutiez, J. Seguin, M. Fonvielle, P. Belin, I. B. Jacques, E. Favry, M. Arthur and M. Gondry, *Nucleic Acids Res.*, 2014, **42**, 7247-7258.
32. G. Bourgeois, J. Seguin, M. Babin, P. Belin, M. Moutiez, Y. Mechulam, M. Gondry and E. Schmitt, *J Struct Biol*, 2018, **203**, 17-26.
33. M. Gondry, I. B. Jacques, R. Thai, M. Babin, N. Canu, J. Seguin, P. Belin, J. L. Pernodet and M. Moutiez, *Front Microbiol*, 2018, **9**, 46.
34. S. Meng, W. Han, J. Zhao, X. H. Jian, H. X. Pan and G. L. Tang, *Angew Chem Int Ed Engl*, 2018, **57**, 719-723.
35. J. Liu, H. Yu and S. M. Li, *Appl Microbiol Biotechnol*, 2018, **102**, 4435-4444.
36. T. W. Giessen, A. M. von Tesmar and M. A. Marahiel, *Chem Biol*, 2013, **20**, 828-838.
37. T. W. Giessen, A. M. von Tesmar and M. A. Marahiel, *Biochemistry*, 2013, **52**, 4274-4283.
38. I. B. Jacques, M. Moutiez, J. Witwinowski, E. Darbon, C. Martel, J. Seguin, E. Favry, R. Thai, A. Lecoq, S. Dubois, J. L. Pernodet, M. Gondry and P. Belin, *Nat. Chem. Biol.*, 2015, **11**, 721-727.
39. J. Seguin, M. Moutiez, Y. Li, P. Belin, A. Lecoq, M. Fonvielle, J. B. Charbonnier, J. L. Pernodet and M. Gondry, *Chem Biol*, 2011, **18**, 1362-1368.
40. E. D. James, B. Knuckley, N. Alqahtani, S. Porwal, J. Ban, J. A. Karty, R. Viswanathan and A. L. Lane, *ACS Synth. Biol.*, 2016, **5**, 547-553.
41. H. Yu, X. Xie and S. M. Li, *Org Lett*, 2018, **20**, 4921-4925.
42. J. B. Patteson, W. Cai, R. A. Johnson, K. C. Santa Maria and B. Li, *Biochemistry*, 2018, **57**, 61-65.
43. K. Brockmeyer and S.-M. Li, *J. Nat. Prod.*, 2017, **80**, 2917-2922.
44. N. M. Vior, R. Lacret, G. Chandra, S. Dorai-Raj, M. Trick and A. W. Truman, *Appl Environ Microbiol*, 2018, **84**, e02828-02817.
45. N. Alqahtani, S. K. Porwal, E. D. James, D. M. Bis, J. A. Karty, A. L. Lane and R. Viswanathan, *Org. Biomol. Chem.*, 2015, **13**, 7177-7192.
46. N. Canu, P. Belin, R. Thai, I. Correia, O. Lequin, J. Seguin, M. Moutiez and M. Gondry, *Angew Chem Int Ed Engl*, 2018, **57**, 3118-3122.
47. M. A. Skinnider, C. W. Johnston, N. J. Merwin, C. A. Dejong and N. A. Magarvey, *BMC Genom*, 2018, **19**, 1-11.
48. M. Gondry, S. Lautru, G. Fusai, G. Meunier, A. Ménez and R. Genet, *Eur. J. Biochem.*, 2001, **268**, 1712-1721.
49. S. Lautru, M. Gondry, R. Genet and J.-L. Pernodet, *Chem Biol*, 2002, **9**, 1355-1364.
50. P. Belin, M. H. Le Du, A. Fielding, O. Lequin, M. Jacquet, J.-B. Charbonnier, A. Lecoq, R. Thai, M. Courçon, C. Masson, C. Dugave, R. Genet, J.-L. Pernodet and M. Gondry, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 7426-7431.
51. K. Dornevil, I. Davis, A. J. Fielding, J. R. Terrell, L. Ma and A. Liu, *J Biol Chem*, 2017, **292**, 13645-13657.
52. W. Tian, C. Sun, M. Zheng, J. R. Harmer, M. Yu, Y. Zhang, H. Peng, D. Zhu, Z. Deng, S. L. Chen, M. Mobli, X. Jia and X. Qu, *Nat Comm*, 2018, **9**, 4428.
53. T. Yao, J. Liu, Z. Liu, T. Li, H. Li, Q. Che, T. Zhu, D. Li, Q. Gu and W. Li, *Nat Comm*, 2018, **9**, 1-13.
54. T. Robbins, Y. C. Liu, D. E. Cane and C. Khosla, *Curr Opin Struct Biol*, 2016, **41**, 10-18.
55. M. J. Cryle, S. G. Bell and I. Schlichting, *Biochemistry*, 2010, **49**, 7282-7296.
56. K. Fukushima, K. Yazawa and T. Arai, *J Antibiot*, 1973, **26**, 175-176.
57. R. L. Uffen and E. Canale-Parola, *J Bacteriol*, 1972, **111**, 86-93.
58. A. J. Kluyver, J. P. van der Walt and A. J. van Triet, *Proc. Natl. Acad. Sci. U. S. A.*, 1953, **39**, 583-593.
59. M. R. Tang, D. Sternberg, R. K. Behr, A. Sloma and R. M. Berka, *Ind. Biotechnol.*, 2006, **2**, 66-74.
60. D. J. Krause, J. Kominek, D. A. Opulente, X.-X. Shen, X. Zhou, Q. K. Langdon, J. DeVirgilio, A. B. Hulfachor, C. P. Kurtzman, A. Rokas and C. T. Hittinger, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 11030-11035.
61. K. J. McLean, P. Carroll, D. G. Lewis, A. J. Dunford, H. E. Seward, R. Neeli, M. R. Cheesman, L. Marsollier, P. Douglas, W. E. Smith, I. Rosenkrands, S. T. Cole, D. Leys, T. Parish and A. W. Munro, *J Biol Chem*, 2008, **283**, 33406-33416.
62. M. Fonvielle, M. H. Le Du, O. Lequin, A. Lecoq, M. Jacquet, R. Thai, S. Dubois, G. Grach, M. Gondry and P. Belin, *J Biol Chem*, 2013, **288**, 17347-17359.
63. V. G. Dumas, L. A. Defelipe, A. A. Petruk, A. G. Turjanski and M. A. Marti, *Proteins*, 2013, **82**, 1004-1021.
64. M. E. Kavanagh, A. G. Coyne, K. J. McLean, G. G. James, C. W. Levy, L. B. Marino, L. P. S. de Carvalho, D. S. H. Chan, S. A. Hudson, S. Surade, D. Leys, A. W. Munro and C. Abell, *J. Med. Chem.*, 2016, **59**, 3272-3302.
65. J. Kim, J. A. Ashenhurst and M. Movassaghi, *Science*, 2009, **324**, 238-241.
66. R. Raju, A. M. Piggott, M. Conte, W. G. Aalbersberg, K. Feussner and R. J. Capon, *Org Lett*, 2009, **11**, 3862-3865.
67. L. M. Alkhalaf and K. S. Ryan, *Chem Biol*, 2015, **22**, 317-328.
68. T. Kamiya, S. Maeno, M. Hashimoto and Y. Mine, *J Antibiot* 1972, **25**, 576-581.
69. A. Zwiefka, H. Kohn and W. R. Widger, *Biochemistry*, 1993, **32**, 3564-3570.
70. M. Iseki, T. Miyoshi, T. Konomi and H. Imanaka, *J Antibiot* 1980, **33**, 488-493.
71. D. K. Liscombe, G. V. Louie and J. P. Noel, *Nat Prod Rep*, 2012, **29**, 1238-1250.
72. P. Fu, P. Liu, H. Qu, Y. Wang, D. Chen, H. Wang, J. Li and W. Zhu, *J Nat Prod*, 2011, **74**, 2219-2223.
73. Q. Zhang, S. Li, Y. Chen, X. Tian, H. Zhang, G. Zhang, Y. Zhu, S. Zhang, W. Zhang and C. Zhang, *J Antibiot* 2013, **66**, 31-36.
74. M. E. Tanner, *Nat Prod Rep*, 2015, **32**, 88-101.