Unique marine derived cyanobacterial biosynthetic genes for chemical diversity

Journal:	*Natural Product Reports*
Manuscript ID	NP-REV-08-2015-000097.R1
Article Type:	Review Article
Date Submitted by the Author:	14-Nov-2015

Complete List of Authors:	Kleigrewe, Karin; Scripps Institution of Oceanography, University of California San Diego Gerwick, Lena; Scripps Institution of Oceanography, Center for Marine Biotechnology and Biomedicine Sherman, David; University of Michigan, Life Sciences Institute, and Departments of Medicinal Chemistry, Chemistry and Microbiology & Immunology Gerwick, William; Scripps Institution of Oceanography, Center for Marine Biotechnology and Biomedicine

Journal Name

ARTICLE

Received 00th January 20xx,

Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Unique marine derived cyanobacterial biosynthetic genes for chemical diversity

Karin Kleigrewe[a*], Lena Gerwick[a], David H Sherman[c], William H. Gerwick[a,c]

Cyanobacteria are a prolific source of structurally unique and biologically active natural products that derive from intriguing biochemical pathways. Advancements in genome sequencing have accelerated the identification of unique modular biosynthetic gene clusters in cyanobacteria and reveal a wealth of unusual enzymatic reactions involved in their construction. This article examines several interesting mechanistic transformations involved in cyanobacterial secondary metabolite biosynthesis with a particular focus on marine derived modular polyketide synthases (PKS), nonribosomal peptide synthetases (NRPS) and combinations thereof to form hybrid natural products. Further, we focus on the cyanobacterial genus *Moorea* and the co-evolution of its enzyme cassettes that create metabolic diversity. Progress in the development of heterologous expression systems for cyanobacterial gene clusters along with chemoenzymatic synthesis makes it possible to create new analogs. Additionally, phylum-wide genome sequencing projects have enhanced the discovery rate of new natural products and their distinctive enzymatic reactions. Summarizing, cyanobacterial biosynthetic gene clusters encode for a large toolbox of novel enzymes that catalyze unique chemical reactions, some of which may be useful in synthetic biology.

[a] *Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, USA; *current affiliation: Bavarian Center for Biomolecular Mass Spectrometry (BayBioMS), Chair of Food Chemistry and Molecular Sensory Science, Technische Universität München, Freising, Germany*
[b] *Life Sciences Institute, University of Michigan, Ann Arbor, Michigan, USA*
[c] *Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, USA*

1. Introduction

Cyanobacteria are a prolific source of bioactive natural products.[1] They are among the oldest life forms on Earth,[2] and thus have a long evolutionary history that has allowed their adaption to a multiplicity of habitats like freshwater ponds and lakes, oceans, deserts, thermal springs, rocks, polar ice and even animals and plants.[3] Moreover, cyanobacterial natural products are a prolific source of new bioactive compounds with high potential as future drug leads.[4, 5] Nevertheless, some of these compounds can be harmful to humans; for example, in 2014 in Toledo, Ohio, freshwater reservoirs became contaminated by cyanobacterial blooms that produced the hepatotoxic microcystins.[6]

This review focuses mainly on cyanobacteria from the marine environment as they are exceptionally rich in producing unique metabolites that have promising applications in the areas of

anticancer, antibiotic and anti-inflammatory therapeutics.[7] These compounds are often produced by polyketide synthase (PKS) or nonribosomal peptide synthetase (NRPS) assembly lines which are subsequently transformed post-assembly to modify the initial molecular scaffold.[8] Curacin A (Figure 1), a compound with potent anti-cancer properties, is an excellent example of a natural product rich in diverse functional groups.[9] Its chemical structure bears a cyclopropyl ring, a *cis*-double bond, a thiazoline ring and a terminal alkene.[10] Understanding how this compound is biochemically synthesized gives insight into the unique enzymatic machinery of cyanobacteria. This review also summarizes current examples of the co-evolution of enzymes, which involves highly similar genetic architectures, but results in diverse functional groups.[11, 12] Understanding the enzymatic mechanisms behind this unique chemistry may aid in the creation of novel chemistry through heterologous expression or chemo-enzymatic synthesis of novel compounds. In this context, phylum-wide genome sequencing projects are helping to uncover novel biosynthetic gene clusters.[13] Lessons learned from Nature will guide future engineering efforts aimed at creating bio-inspired secondary metabolites *in vitro* and *in vivo*. This review covers the literature from 2010 until July 2015. A previous review from Jones et al. covered these topics up to 2010.[14]

2. Co-evolution of enzymes for metabolic diversification

Genome sequencing projects are rapidly enhancing the number of available genomes to mine for biosynthetic pathways.[13] Recent biosynthetic pathway comparisons have demonstrated the evolutionary relatedness between specific biosynthetic enzyme cassettes. This section highlights cases in which highly similar genetic architectures have resulted in different chemical functional groups within the *Moorea* species.

2.1. β-branching modules in the apratoxin, curacin, and jamaicamide pathways

The curacin A, jamaicamide A and apratoxin A biosynthetic pathways are prominent examples of hybrid PKS/NRPS systems with unique β-branching modules.[11] These natural products were isolated from *Moorea producens* 3L (formerly known as *Lyngbya majuscula*[15]), *Moorea producens* JHB and *Moorea bouillonii* PNG, respectively. All three compounds have different biological activities; curacin A is a potent inhibitor of tubulin polymerization[9], jamaicamide A is a sodium channel blocker[16] and apratoxin A is a highly cytotoxic compound causing G_1 cell cycle arrest and apoptosis.[17,18] All three pathways consist of a tandem $ACP_{2/3}$ (ACP_{I}-ACP_{II}-(ACP_{III})) di-or tridomain (see 3.1), a discrete ACP_{IV}, a 3-hydroxy-3-methylglutaryl CoA-synthase enzyme (HCS), a dehydratase (ECH_1), a decarboxylase (ECH_2) and an enoyl reductase (ER) domain (Figure 1).[11, 12, 16] Additionally, the curacin A and jamaicamide pathways include a non-heme Fe(II), α-ketoglutarate-dependent halogenase (CurA, JamE)[11] (see 3.4.1) whereas the apratoxin pathway lacks one of the tandem ACPs associated with the HCS cassette.[12]

Even though these enzyme cassettes are quite similar, they are responsible for the production of very different functional groups. In the curacin A pathway these enzymes produce a cyclopropane ring, in the jamaicamide pathway a vinyl chloride and in the apratoxin pathway a secondary *β*-methyl group (Figure 1). The catalytic domains and discrete enzymes for creating these functional groups span from CurA to CurF, from JamE to JamJ and from AprB to AprG. These enzyme cassettes share an extraordinarily high sequence identity (75.7% - 95.3%, Figure 1).[11]

Gu *et al.* overexpressed and purified individual enzymes and performed comparative biochemical studies on the variant functions and selectivity using the curacin A and jamaicamide A biosynthetic components.[11] At that time, the apratoxin pathway was not yet described.[12] Initial analysis of the curacin pathway suggested that *cur*D encodes for a hydroxy methyl glutaryl (HMG)-CoA synthase-like gene (HCS) that catalyzes the formation of HMG-ACP from an ACP_{IV}-bound acetate and acetoacetyl-ACP_{I-III}.[19] Next, the ECH_1 enzyme was proposed to function as a dehydratase to create 3-methylglutaconyl-ACP, which is subsequently decarboxylated by the ECH_2 enzyme to form 3-methyl-crotonyl-ACP.[11, 20, 21] Initial exploration of how the HMG-ACP intermediate is converted into the 2-methylcyclopropane unit of curacin A focused on the enoylreductase (ER) in combination with different substrates.[20] When the HMG-ACP substrate was incubated with CurE ECH_1 and CurF ECH_2, the expected 3-methylcrotonyl-ACP product was produced. Upon addition of the enoylreductase (ER) CurF and NADPH, the product was reduced to provide a secondary methyl group, as observed in apratoxin A (Figure 1).[11]

However, a halogenase (Hal) is also required in the formation of the cyclopropyl-ring of curacin A as well as the vinyl-chloride group of jamaicamide A.[11] A non-heme Fe(II), α-ketoglutarate-dependent halogenase is present in both of the curacin A and jamaicamide A biosynthetic gene clusters. Biochemical studies showed that the Cur Hal specifically forms γ-Cl-1-ACP_{II}.[11] Next, γ-Cl-1-ACP_{II} reacts sequentially with Cur ECH_1 and ECH_2 to form (*E*)-4-chloro-3-methylbut-2-enyl-ACP. Upon addition of the Cur ER, the cyclopropyl ring was formed (Figure 1).[11] To understand the formation of the vinyl-chloride group present in jamaicamide A, Jam ECH_1, Jam ECH_2 and Jam ER were incubated with suitable, unnatural substrates derived from the Cur pathway. It was observed that Jam ECH_2 decarboxylates HMG γ-Cl-2-ACP_{II} to yield the corresponding β,γ enoyl thioester instead of the α,β enoyl thioester.[11] In addition Jam ECH_2 showed high regiochemical control generating exclusively the β,γ double bond with 85% in the *E* configuration.[11] However, because the canonical Jam ER catalyzes the reduction of an α,β enoyl thioester, its function is superfluous in the formation of jamaicamide.[11] In summary, the HMG cassettes in the curacin A, apratoxin A and jamaicamide A pathways share a close ancestral relationship, yet demonstrate a remarkable degree of evolutionary diversification in natural product biosynthesis. In this case, small modifications in the genomic sequence have an immense impact on the resulting chemical structures and their associated biological activities.

2.2 A regulatory enzyme associated with the curacin and jamaicamide biosynthetic pathway leads to the discovery of the columbamides and their biosynthetic pathway

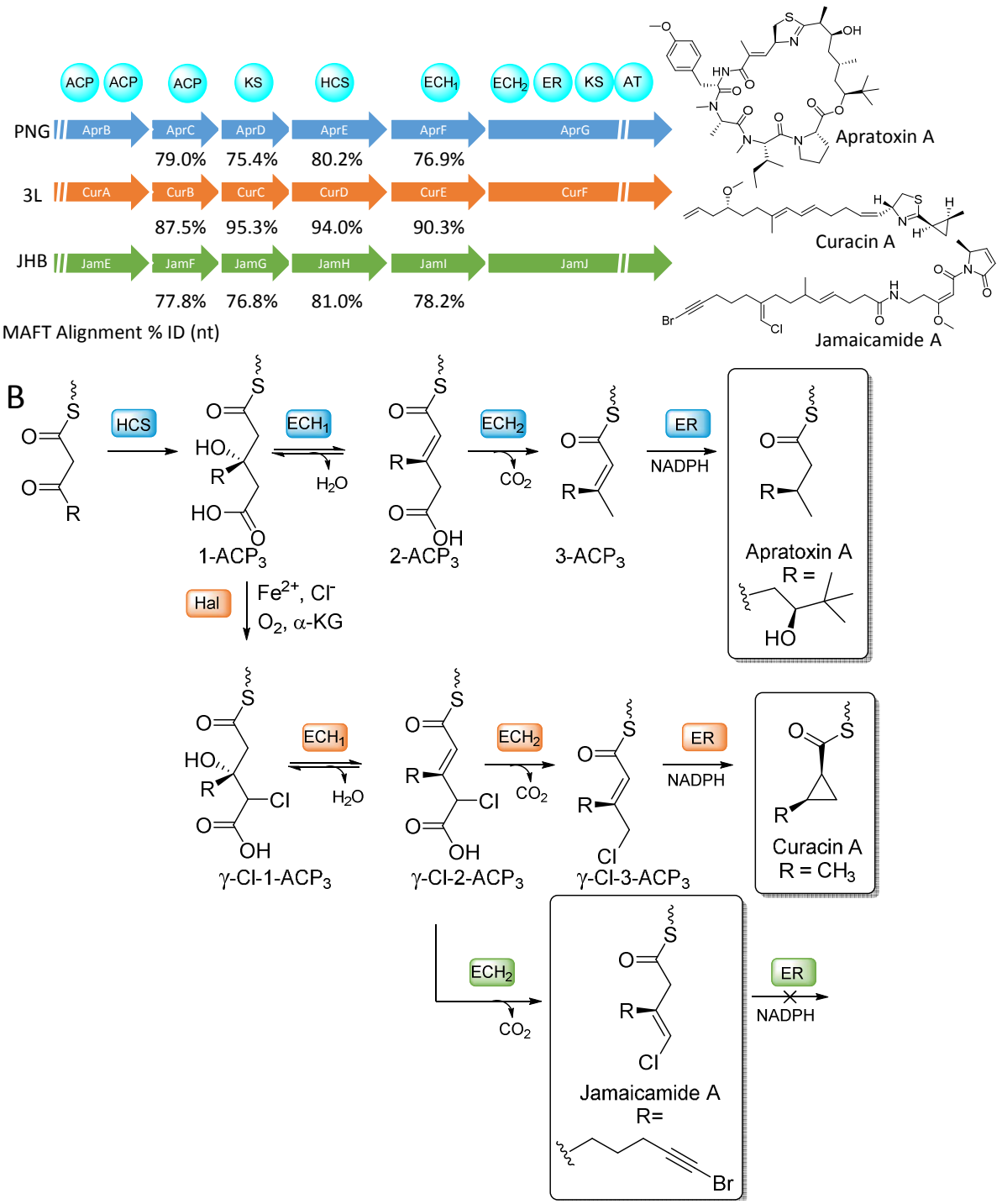


Figure 1: A: Comparison of the apratoxin A, curacin A, and jamaicamide A biosynthetic pathways and their percentages sequence similarity (bottom row of numbers compare jamaicamide A and apratoxin A). B. Evolution of β-branching enzymes in the apratoxin, curacin and jamaicamide biosynthetic pathways involves a functional diversification of enoyl reductases (ERs in Apr, Cur) and alternative regiochemical control by enoyl CoA hydratases (functionally a decarboxylase, ECH2 in Apr, Cur, Jam) (modified after Gu et al.)[11], ACP: acyl carrier protein, KS: β-ketoacyl-ACP Synthase, HCS: 3-hydroxy-3-methylglutaryl CoA synthase, ECH1: dehydratase, ECH2: decarboxylase, ER: enoyl reductase, AT: acyl transferase, Hal: halogenase.

The columbamides were discovered by combining new mass spectrometric tools and molecular networking with genome mining approaches.[22] The new metabolites are a novel class of di- and tri-chlorinated acyl-amides that exhibit relatively potent cannibinomimetic activity.[22] Genome comparisons and analyses revealed a nearly identical genomic region, identified by BLAST analysis as a putative regulatory serine histidine kinase, that was located in association with both the curacin A and jamaicamide A

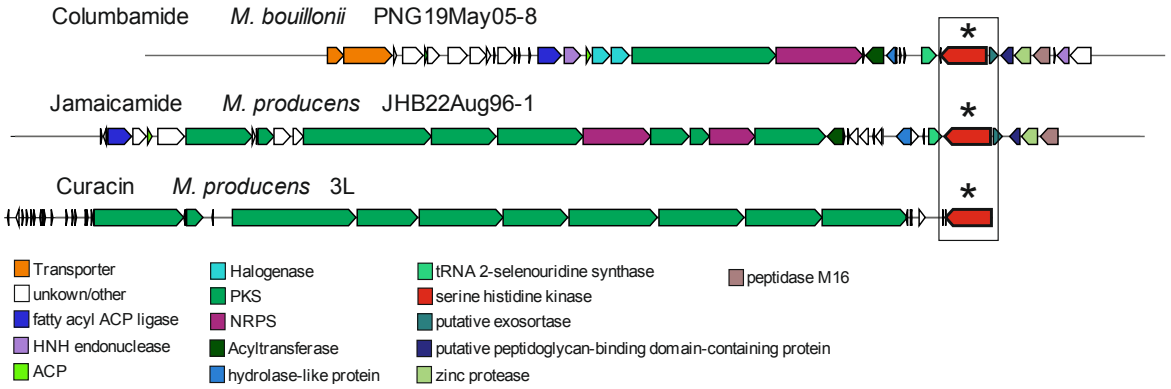


Figure 2: MultiGeneBlast result for the columbamide, jamaicamide A and curacin A biosynthetic gene clusters. The 3' end of each pathway contains an open reading frame for a putative regulatory serine histidine kinase (*).

biosynthetic gene clusters.[16, 19] A highly similar gene (Figure 2) was also found in the genome of *Moorea bouillonii* PNG, however, the product of an adjacent biosynthetic pathway was unknown.[22] Recognition of key structural features implied by this genomic information, such as overall molecular size and inclusion of both hetero and halogen atoms, was compared with the LC-MS/MS fragmentation-based Molecular Networks, and HR-MS datasets.[22-24] It was thus possible to link this gene cluster to the expressed metabolites and *vice versa*.

Figure 3 displays the proposed biosynthetic pathway for the columbamides.[22] ColA initiates the pathway by loading dodecanoic acid onto the acyl carrier protein ColC. ColB, a presumed HNH endonuclease, may be involved in horizontal gene transfer of a part or the entirety of the cluster, and does not appear to have a functional biosynthetic role. Next, ColD and ColE are most likely responsible for multiple halogenations of the acyl chain at the ω-terminus and ω-7 positions. These enzymes share sequence identity to *p*-aminobenzoate *N*-oxygenases (ColD: 27% identity, 47% similarity, ColE: 30% identity, 44% similarity), to the protein CylC (ColD: 48% identity, 66% similarity, ColE: 47% identity, 62% similarity) which is most likely a cryptic halogenase involved in carbon-carbon bond formation during cylindrocyclophane biosynthesis, and to the recently described BrtJ enzyme (ColD: 43% identity, 60% similarity, ColE: 44% identity, 57% similarity) which is involved in the formation of the bartolosides, compounds with a chlorinated dialkylresorcinol core structure.[22, 25-27] It may be that the mechanism of halogenation at unactivated carbon atoms on alkyl chains is similar to the oxidation of an aminoarene to a nitroarene.[22, 27] Next, a bimodular PKS motif extends the halogenated acyl ACP by two rounds of ketide extension. A DH domain is expected in both of these PKS extension modules, but is missing from the former and only present in the second module. Similar to a missing dehydratase domain in the curacin A cluster (see 3.2), we speculate that the downstream DH serves this function for both extensions.[22, 28] Subsequently, the NRPS module ColG extends the acyl ACP via an amide linkage to a serine residue, and this is methylated at both its hydroxy group and amide nitrogen atom via *O*- and *N*-methyltransferases. Finally, the peptidyl carrier protein bound product is released by a reductase mechanism (48% identity, 66% similarity to the release domain of the lyngbyatoxin biosynthetic gene cluster Ltx) as a primary alcohol, and this is acetylated, possibly by ColI, to form columbamide A and B.[22, 29] This presumed regulatory serine histidine kinase found associated with these three relatively highly expressed pathways could be used in other genome mining projects to identify additional natural product biosynthetic pathways with high constitutive expression.[22]

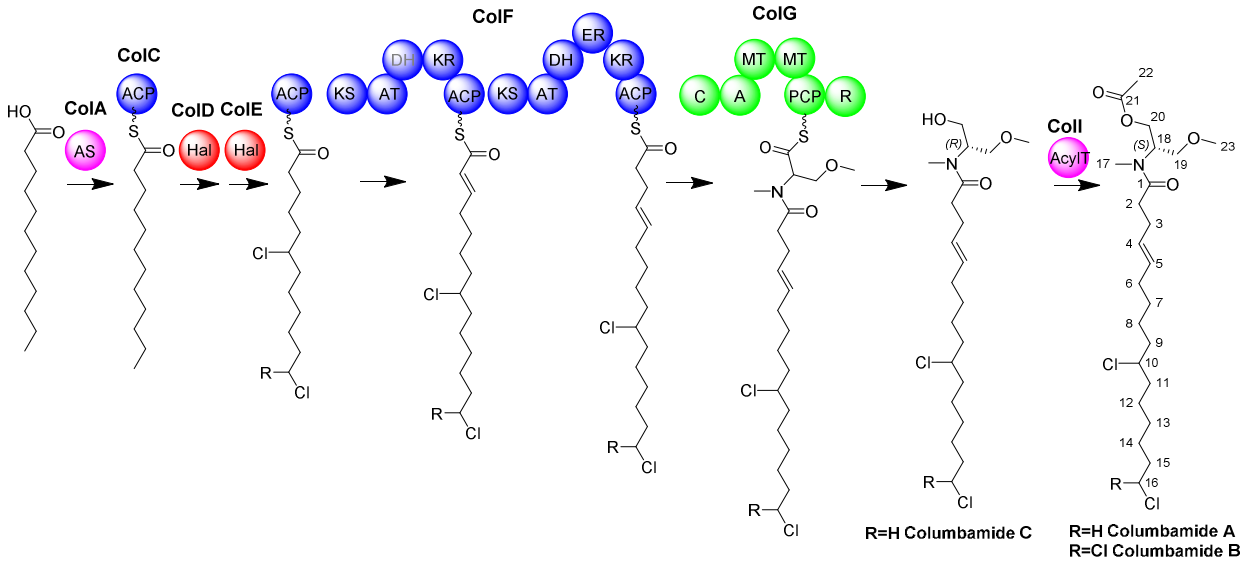


Figure 3: Proposed biosynthetic pathway for the columbamides produced by M. bouillonii PNG. AS: acyl synthetase, ACP: acyl carrier protein, Hal: halogenase, KS: β-ketoacyl-ACP synthase; AT: acyl transferase, DH: β-hydroxy-acyl-ACP dehydratase, KR: β-ketoacyl-ACP reductase, ER: enoyl reductase, C: condensation domain, A: adenylation domain, MT: methyltransferase, PCP: peptidyl carrier protein, R: Reductase, AcylT: Acyltransferase. The first DH domain in ColF is not detectable through bioinformatics, therefore it is coloured in light grey. As discussed in the text, the second DH domain of ColF presumably catalyzes both dehydratase reactions.[22]

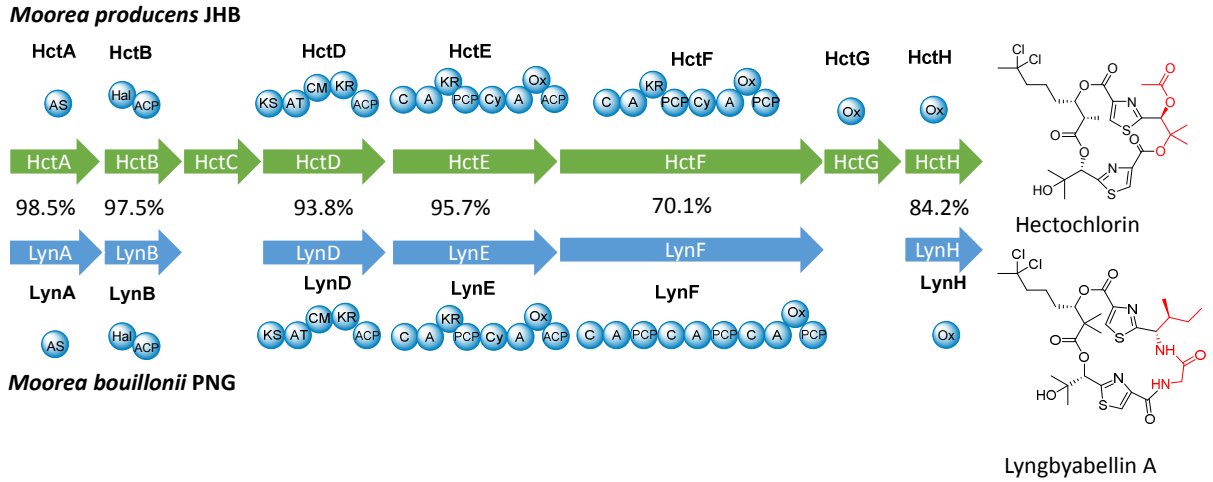


Figure 4: Comparison of the hectochlorin and lyngbyabellin A biosynthetic pathways. The percentages show the MAFFT alignment % ID (nt). Portions of structures in black are identical whereas in red are sections differing between the two molecules. AS: acyl synthetase, Hal: Halogenase, ACP: acyl carrier protein, KS: β-ketoacyl-ACP synthase, AT: acyl transferase, CM: C-methyltransferase, KR: β-ketoacyl-ACP reductase, C: condensation domain, A: adenylation domain, PCP: peptidyl carrier protein, Cy: cyclization domain, Ox: oxidation domain.

2.3 Hectochlorin and lyngbyabellin: Related biosynthetic pathways in different *Moorea* strains

Remarkably, *Moorea producens* JHB, which was collected in Hector's Bay Jamaica[30] and *Moorea bouillonii* PNG, which was collected from Pigeon Island, Papua New Guinea[31] share an almost identical 50 Kb hybrid NRPS/PKS biosynthetic pathway.[32] Both hectochlorin A[33] and lyngbyabellin A[34] possess antifungal and cytotoxic properties due to their capacity to induce the hyperpolymerization of actin filaments.[30] A targeted gene cloning/sequencing approach was used to characterize the hectochlorin biosynthetic cluster whereas a genome sequencing effort led to characterization of the lyngbyabellin pathway.[32, 33] The initial acyl-ACP synthetase (HctA, LynA) activates hexanoic acid which is then halogenated by HctB/LynB to form the gem-dichloro group at the penultimate carbon atom (Figure 4, see 3.4.2). This is followed by a single acetate extension along with a C-methylation event (HctD/LynD).[33] Interestingly, in the case of lyngbyabellin A, a double C-methylation occurs at this juncture whereas a single methylation is observed in hectochlorin A. The biochemical basis for this double methylation from a single C-methyl transferase domain remains unknown. Next, the NRPS module (C-A-KR-PCP-Cy-A-Ox-ACP) (HctE/LynE) incorporates the α-hydroxy acid 2,3-dihydroxyisovaleric acid. This is followed by the addition of a cysteine residue that becomes heterocyclized and oxidized to form a thiazole ring.[32, 33] Four of the first five genes (Hct A, B, D, E/Lyn A, B, D, E) of these two pathways share a sequence identity of 93.8%-98.5% whereas HctC, a transposase of unknown function, is missing in the *M. bouillonii* PNG cluster. Figure 4 shows the chemical structure of hectochlorin and lyngbyabellin A. The black portions of the two structures are formed by the above-mentioned proteins HctA/LynA, HctB/LynB, HctD/LynD, HctE/LynE.

The two pathways diverge at this point in the biosynthetic sequence. In hectochlorin, the bimodular HctF gene, whose domain architecture is highly similar to that of HctE, codes for an NRPS module which incorporates a second α-keto acid, following its reduction to 2,3-dihydroxyisovaleric acid, and a cysteine residue, the latter again becoming cyclized and oxidized to a thiazole residue.[33] A candidate gene encoding a protein to introduce the acetate group at C-14 is lacking, and might occur as a post-assembly modification from a protein encoded outside of the gene cluster.[33] In contrast, LynF of the lyngbyabellin A pathway is a trimodular NRPS which incorporates glycine, isoleucine and cysteine, the latter becoming cyclized to a thiazole. Following incorporation of the latter cysteine residue in both pathways, the two molecules are released from the megasynthases with concurrent macrocyclization (Figure 4).[32] Lastly, HctG and HctH in the hectochlorin pathway, and LynH in the lyngbyabellin pathway, are putative P450 monooxygenases that may be responsible for oxidation events during their biosynthesis.[33] It is remarkable that such similar biosynthetic pathways are present in two cyanobacteria from such distant locations, both in their overall architecture and nucleotide sequence, and is likely an example of horizontal gene transfer with ensuing neofunctionalization; future phylogenetic studies may shed light on this phenomenon. Recent genome sequencing efforts have demonstrated that a highly similar pathway is present in another *Moorea* strain collected from Strawn Island, Palmyra Atoll; however, the product of the pathway has not yet been identified (unpublished data).

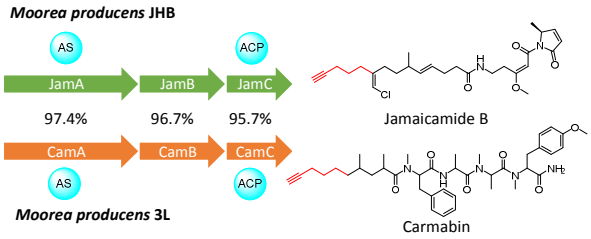


Figure 5: Nearly identical acyl-ACP loading domains for the jamaicamide and carmabin biosynthetic pathways. The percentages show the MAFFT alignment % ID (nt). Colored in red is the same substructure in each. AS: acyl synthetase, ACP: acyl carrier protein.

2.4 Carmabin and jamaicamide A: Closely related acyl-ACP loading domains in two divergent biosynthetic pathways

The first steps in the biosynthesis of the sodium-channel blocker jamaicamide A,[16] which is produced by *Moorea producens* JHB, and

the antimalarial compound carmabin A,[35] which is produced by *M. producens* 3L, incorporate the identical hexynoic acid moiety.[35, 36] Comparison of their respective biosynthetic gene clusters has revealed that the first three genes (Jam A-C, Cam A-C) are highly conserved with a sequence identity of 95.7-97.4% (Figure 5).[16, 37] Mining the publically available microbial genomes for this *jam* A-C motif identifies an additional 80 gene clusters encoding close homologs.[38] The mechanism behind the alkyne formation was recently studied by Zhu and colleagues[38] with the goal of tagging natural products with a unique chemical handle so that they could be visualized and assist in determining their mode of action.[38] Jam A is a fatty acid-acyl ligase with preference for 5-hexanoic acid or 5-hexynoic acid, whereas Jam B was characterized as a membrane-associated fatty acid desaturase that generates the terminal alkyne from hexenoic acid loaded onto the JamC acyl carrier protein.[16, 38] JamB has a strict substrate specificity for the acyl group (hexanoyl-JamC, 5-hexenoyl-JamC) and the acyl carrier protein (JamC, CamC). Therefore, only 5-hexynoyl is incorporated into alkyne-tagged natural products using the JamA-C *in situ* system.[38] This system was successfully used in the production of HsPKS1, a tagged product deriving from a type III PKS found in the plant *Huperzia serrate*, as well as in forming a tagged antimycin analog (Figure 6).[38] With their newly incorporated alkyne functionalities, these molecules are suitable for further derivatization using Click chemistry. The combination of cyanobacterial genome mining with such biochemical and genetic manipulations is an effective approach to develop new chemoenzymatic tools and cellular probes.

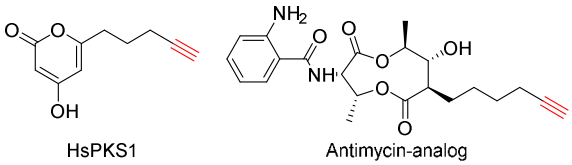


Figure 6: Terminal alkyne functionalities were genetically added to HsPKS1 and antimycin, demonstrating the utility of such genetic manipulations to create cellular probes.[38]

3. Unique mechanistic transformations in cyanobacterial pathways

Cyanobacteria are well known for producing natural products with a wide range of unique structural features.[1, 39, 40] Advances in protein crystallography together with NMR titration and *in vitro* expression of proteins is shedding light on the fascinating enzymatic machinery behind the formation of some of these functional groups.[41] The following section summarizes recent findings on topics such as the function of tandem ACPs in β-branching cassettes and cryptic halogenation during curacin A biosynthesis, and discusses their potential future applications in the chemoenzymatic synthesis of novel bioactive compounds.

3.1 β-branching: The role of the tandem acyl carrier proteins

Tandem tri- or di-ACP domains[43] are frequently associated with the hydroxymethylglutaryl (HMG) producing enzyme cassettes in hybrid PKS biosynthetic pathways; for example, these are observed in the curacin A[19], jamaicamide,[16] apratoxin,[9] bacillaene,[44] mupirocin[45] and pederin pathways.[46] They catalyze consecutive modifications on tandem ACP-linked substrates, and have been demonstrated to improve product yields in the mupirocin and polyunsaturated fatty acid pathways.[47, 48] A recent study overexpressed and purified the tandem CurA ACP_3 domains in different constellations.[42]

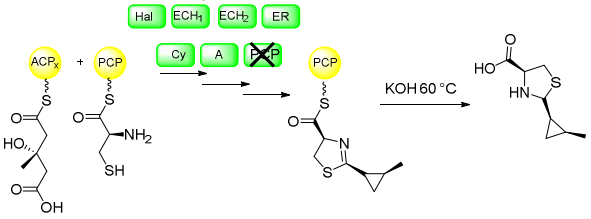


Figure 7: One-pot reactions with variant HMG-linked tandem ACP_3 constructs, truncations and mutants to monitor the product yield.[42] ACP: acyl carrier protein, PCP: peptidyl carrier protein Hal: Halogenase, ECH_1: dehydratase, ECH_2: decarboxylase, ER: enoyl reductase, Cy: cyclization domain, A: adenylation domain.

N-terminal His-tagged fusion proteins with the CurA ACP_I-ACP_{II}-ACP_{III}-CD tetradomain and ACP_I-ACP_{II}-ACP_{III} tridomain were prepared, as well as five truncation constructs, including ACP_I-ACP_{II}, ACP_{II}-ACP_{III}, ACP_I, ACP_{II}, and ACP_{III} and single and double Ser →Ala ACP active-site mutations on the ACP_I-ACP_{II}-ACP_{III} and ACP_I-ACP_{II}-ACP_{III}-CD constructs.[42] The "CD" section is a small C-terminal domain, which flanks the triple-ACP. *In vitro* experiments with ACP-linked HMG and PCP-linked L-cysteine substrates and the enzymes Hal, ECH_1, ECH_2, ER and a Ser → Ala mutant Cy-A-PCP (Figure 7) demonstrated that single ACP domains provides only 10% of the yield compared to the complete ACP_I-ACP_{II}-ACP_{III}-CD construct. The tridomain without the CD-portion produced a 37% yield, suggesting a weak cooperation between this motif and the three ACPs. When the CD domain was added to the C-terminus of an ACP, the yield increased from 10% to 25-30% for each single active ACP in the tridomain. The same improvement occurred when two active ACPs were present in the tridomain, underlining the involvement of triple ACPs in product yields.[42] While the single curacin A ACP domains were measured to be close to their calculated molecular weight, purified di- and tridomain ACPs were nearly twice their expected mass as determined by size exclusion chromatography, suggesting an extended globular structure in which the linkers may be flexible and the ACP domains do not interact.[42] Haines et al. used NMR to study the interactions of the β-branching ACPs, but could not observe ACP-ACP synergistic effects as the ACPs appeared to function independently.[49] Consistent with this, other NMR experiments with the triplet CurA ACPs domains did not detect an interaction between ACP_I, ACP_{II} and ACP_{III}.[50] In conclusion, the tandem organization of ACP domains improves the rapid access and pairing of enzymes and substrates and thereby increases product yield. This unique domain structure is therefore of significance for future bioengineering attempts to generate novel natural products in high yields.

3.2 Domain docking for optimized biosynthetic throughput

Biosynthetic pathway fidelity and throughput is dependent on the correct transfer of chain elongation intermediates from one module to the next.[51] Non-covalent C- and N-terminal docking domains promote the protein-protein interaction of upstream ACP and downstream KS domains, and are therefore an attractive target for pathway engineering to produce natural product analogs.[51, 52] Class 2 docking domains (dd) are present in cyanobacterial polyketide synthase pathways,[52] whereas class 1 docking domains have been identified in actinobacterial PKS modules.[51]

In class 1 docking domains, the ^{ACP}dd contains two dimerization helices that make a four-helix-bundle dimer, followed by a C-terminal docking helix, which binds to the coiled-coil formed by the dimerization of the single dd^{KS} helix. Specificity is obtained through the docking interface, which consists of small (~550 Å) complementary hydrophobic surfaces that are surrounded by

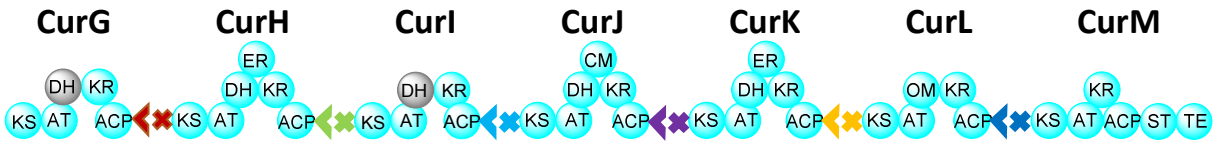


Figure 8: Curacin A docking domains between PKS modules. As discussed in section 3.3, the DH domains coloured in light grey are not detectable by automated bioinformatic analysis. Presumably the following DH domain catalyzes the dehydratase reactions. KS: β-ketoacyl-ACP synthase, AT: acyl transferase, DH: β-hydroxy-acyl-ACP dehydratase, KR: β-ketoacyl-ACP reductase, ACP: acyl carrier protein, ER: enoyl reductase, CM: C-methyltransferase, OM: O-methyltransferase, ST: sulfotransferase, TE: thioesterase.

electrostatic interactions. The ^{ACP}dd docking helices and the dd^{KS} coiled-coil helices are roughly parallel, with the linkage to the upstream ACP directed away from the downstream KS. Long linkers (30-50 amino acids) between the ACP and the ^{ACP}dd promote movement of the carrier protein to the downstream KS and the catalytic domains in the upstream module.[51]

Cyanobacterial class 2 docking domains are able to direct the upstream ACP toward the downstream KS directly.[52] In addition, the class 2 ^{ACP}dds are ~40 amino acids shorter than the class 1 ^{ACP}dds and are thus too short to have the dimerization region. Class 1 and 2 dd^{KS}s are similar in length, but the N-terminal region of the class 2 dd^{KS}s is too polar to form a coiled-coil.[52] Class 2 docking domains consist of two ^{ACP}dd helices (α1 and α2) and two dd^{KS} helices (αA and αB). The interaction of the ^{ACP}dd α2 and the dd^{KS} αB coiled-coil is similar to the interface that is characteristic for the class 1 docking domains. Moreover, class 2 docking interactions include an additional interaction between ^{ACP}dd α1 and dd^{KS} αA and αB; these are conserved in class 2 docks and extend the docking interface.[52] To test the application for class 2 docking domains in the design of natural product expression systems, the class 1 docking domains of the pikromycin pathway were replaced with the class 2 docking pairs from the curacin A pathway (Figure 8), and resulted in the successful production of pikromycin intermediates.[52] Class 2 docking domains are thus another set of tools that broaden the scope of modules that can be employed for small molecule pathway engineering.

The CurA Hal/ACP$_1$ interaction provides another example of specific interactions between domains. This case was studied by NMR titration experiments and MALDI-based identification of products formed.[50] The Cur Hal distinguishes between the CurA ACP$_1$, CurB ACP and TycB1 PCP with product formation only occurring through interaction with the CurA ACP$_1$. Mutagenesis experiments demonstrated that Cur Hal recognizes an interface consisting mainly of the N-terminus of helix II, the 4'-Ppant arm with its substrate and helix III. Due to the high specificity of Cur Hal for (*S*)-HMG bound to an ACP, the ensuing enzymes in the β-branching cassette, such as ECH$_1$ and ECH$_2$, can be less substrate specific. Indeed, biochemical experiments demonstrated that they can recognize different acyl-CoAs as substrates, albeit with significantly lower efficiency.[50]

3.3 Curacin A dehydratase domains

To understand the substrate specificity of PKS enzymes and the detailed molecular interaction of multidomain modules, crystal structures of dehydratase domains from the curacin A biosynthetic pathway were obtained.[28] In general, the PKS DH domain catalyzes reversible dehydration of the β-hydroxy group thus forming an α,β-double bond, usually in the *trans* but occasionally in the *cis* configuration.[53] The curacin pathway is expected to involve five DH-domain catalyzed reactions (intermediates of the CurG, CurH, CurI, CurJ and CurK catalyzed steps), four of which produce *trans* double bonds and one that forms a *cis* olefin. However, CurG and CurI lack DH domains and CurF has an DH domain despite none being needed at this point in the compound's assembly (Figure 8).[28] The Cur DH domains present in the biosynthetic cluster are moderately similar with sequence identities between 21%-28%, except for those in CurF and CurK, which are 68% identical. Crystal structures of the DH domains in CurF, CurH, CurJ and CurK were obtained.[28] The Cur DHs are similar to dehydratases in eukaryotic fatty acid synthase (FAS). They are dimeric, with the monomers having essentially identical double hotdog folds.[28] Due to the active site location, the PKS intermediate must bend 90° to access the catalytic histidine and aspartate residues. Together with the stereochemistry of the β-hydroxy group, this determines the *cis* or *trans* configuration of the ensuing double bond.[28] The β-hydroxy group configuration depends on the type of KR domain present: KRs of the A-type are predicted to support *cis* double-bond formation by the formation of (3*S*)-3-hydroxyacyl-CoA and KRs of the B-type lead to *trans* double bonds by formation of (3*R*)-3-hydroxyacyl-CoA.[41, 53, 54] Consistent with these predictions, CurH, CurI, CurJ and CurK possess KR domains of the B-type, and these modules form *trans* double bonds.[28] The CurG KR is of the A-type and the CurG intermediate possesses a *cis* double bond. The double bonds formed by the CurG and CurI modules are most likely catalyzed by adjacent modules; the CurF or CurH DH for the CurG product and the CurH or CurJ DH for the CurI product.[28] The mechanism behind this utilization of a DH domain from one module in the operation of another module has not been elucidated, but a similar example is present in the columbamide pathway in *Moorea bouillonii* PNG which lacks a DH domain in the ColF domain.[22, 28]

3.4 Halogenations

3.4.1 New insights into the cryptic halogenation of curacin A

The curacin A pathway halogenase catalyzes a cryptic chlorination reaction which ultimately leads to formation of a cyclopropane ring whereas a very similar halogenase in the jamaicamide pathway catalyzes the formation of vinyl chloride group.[11] The catalytic domains in both halogenases share high sequence identity and belong to the $Fe^{2+}/O_2/\alpha$-ketoglutarate-dependent class of halogenase.[55] These halogenases decarboxylate α-ketoglutarate using dioxygen to form a highly reactive Fe^{4+}-oxo species, which subsequently abstracts a hydrogen atom from the substrate (Figure 9).[55] Subsequently, the substrate radical is quenched by "rebound" halogenation with a chlorine atom. Crystal structures of the CurA halogenase in five distinct ligand states revealed an α-ketoglutarate-mediated conformational switch.[55] The free enzyme, lacking α-ketoglutarate, O_2, Cl^- and (*S*)-HMG-ACP has an open conformation in which the active side lid is disordered. In the open form, loops in the enzyme are collapsed over the (*S*)-HMG-ACP binding site, thereby protecting Fe^{2+} from the (*S*)-HMG-ACP carboxylate, a potentially interfering Fe^{2+} ligand.[55] Binding of α-ketoglutarate

triggers the enzyme to adopt a closed form in which the (*S*)-HMG-ACP binding site is now reachable, and the radical reaction proceeds.[55,56] The halogenations of welwitindolinone (see section 3.8) and hectochlorin (see 3.4.2) follow similar mechanisms.[57, 58]

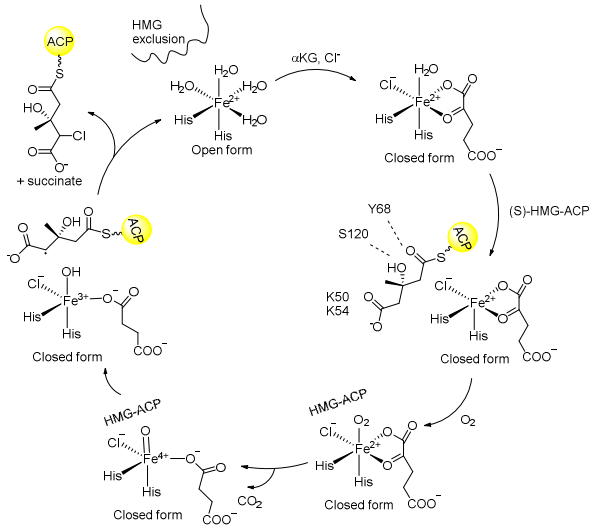


Figure 9: Mechanism of halogenation in the curacin A pathway. The resting enzyme is in the open form and (*S*)-HMG-ACP cannot bind. Binding of α-ketoglutarate and Cl^- trigger the closed form such that (S)-HMG-ACP can bind. Substrate binding permits dioxygen to coordinate with the Fe^{2+} species and leads to oxidative decarboxylation of the α-ketoglutarate intermediate, resulting in formation of the chloro-oxo-ferryl species. This latter intermediate abstracts a hydrogen atom from the substrate and the substrate radical is quenched by chlorination via a "rebound" mechanism.[55, 56]

3.4.2 Hectochlorin HctB: creation of a gem-dichloro functionality

Hectochlorin and the related compounds lyngbyabellin A and dolabellin possess a 5,5-dichlorohexanoyl moiety as an intermediate in their biosynthesis.[33] The halogenase system involved in creating this gem-dichloro functionality introduces these halogen atoms at an unactivated carbon atom. The enzymes involved are an acyl coenzyme A binding protein, a substrate anchoring ACP and the halogenase HctB. Pratter et al. expressed the HctB protein and studied its function *in vitro*, with a focus on the preference of HctB for chlorination over oxygenation.[57, 59] Without significant chloride present, hexanoate and oxohexanoate were formed with small quantities of monochloro- and dichlorohexenoate. Based on their study with different acyl chain lengths, HctB is very specific for metabolizing at the C-5 position, adding either chloride or the oxo moieties at this location. Interestingly, the formation of a vinyl chloride functionality at C-4, C-5 was also observed (Figure 10).[57] It was suggested that the α-ketoglutarate dependent enzyme first transfers either a metal-bound hydroxy or chloro species to C-5, and then in a second reaction the remaining proton at C-5 is abstracted and similarly replaced with a second heteroatom.[57, 59] While this mechanism is similar to that described for CurA halogenase, no other naturally occurring analogs at this position have been detected in extracts of *M. producens* JHB.

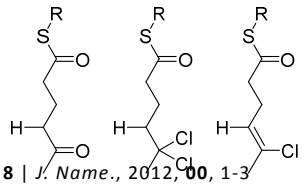


Figure 10: Formation of 5,5-dichloro-hexanoate, 5-oxo-hexanoate or the vinyl-chloride species by reaction of HctB with a hexanoate derivative in vitro.[57]

3.5 Haloalkane dehalogenase DmmA

Bioengineering for industrial applications like biofuels, degradation of environmental pollutants and development of biocatalysts to synthesize bioactive compounds are of great socioeconomic interest. In this regard, a haloalkane dehalogenase (HLDs) DmmA which removes halogens from alkanes by hydrolysis, forming an alcohol, a halide ion and a proton, was obtained from the metagenomic DNA of the marine microbial consortium found in association with *Moorea producens* 3L.[60] Haloalkane halogenases have unique substrate selectivities that cannot be easily predicted by sequence analysis.[60] Nevertheless, the active-site cavity is a reasonable predictor for substrate preferences. DmmA has a large active site and demonstrated activity with bromocyclohexanes, thus suggesting its potential use for chemoenzymatic reactions of large substrates. The catalytic pentad in DmmA is well ordered and consists of Asp 144 as the nucleophile, His315 as the base, Glu168 as the acid and the two halide-stabilizing residues Trp145 and Asn78.[60] The role of DmmA within the marine microbial consortium is unclear but the large active-site cavity suggests that DmmA could be used as a biotechnology tool or reagent.

3.6 Formation of terminal alkenes involving an activating sulfotransferases CurM (ACP-ST-TE) and the OLS ST

Terminal alkenes are found in curacin A as well as in olefins produced by the olefin synthase (OLS) pathway in cyanobacteria.[20, 61, 62, 63] They are formed through sulfotransferases (ST) that catalyze the transfer of a sulfonate group from the sulfonate donor 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to a β-hydroxy group in the substrate (Figure 11).[61, 64, 65] This is followed by action of a unique thioesterase (TE) which acts specifically upon β-sulfated substrates to hydrolyze, decarboxylate and eliminate the sulfate to produce a terminal alkene.[61, 64, 65] Since the sulfonate is subsequently removed through a decarboxylative elimination, these sulfotransferases are known as "activating" STs.[61]

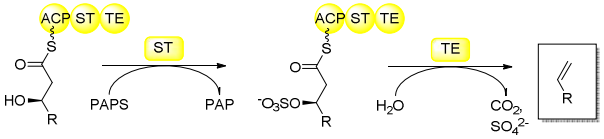


Figure 11: Chain termination using β-hydroxy group activation and decarboxylation in the curacin A and the olefin synthase pathways. The sulfotransferase (ST) sulfonates the β-hydroxy group using 3'-phosphoadenosine 5'-phosphosulfate (PAPS). Subsequently the thioesterase (TE) works in a β-sulfate-dependent manner to hydrolyze, decarboxylate and eliminate sulfate to create the terminal alkene.[61]

This enzyme cassette is of great biotechnological interest as it can be used to produce long-chain hydrocarbons as well as other

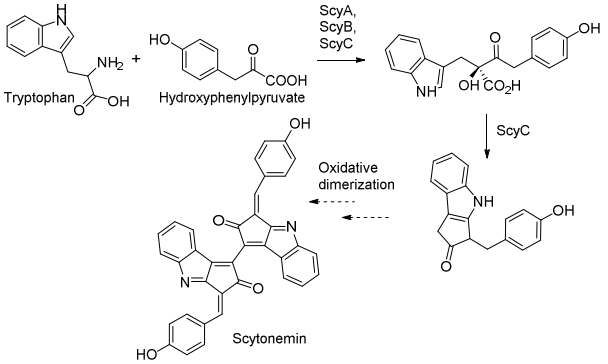


Figure 12: Scytonemin biosynthetic pathway.

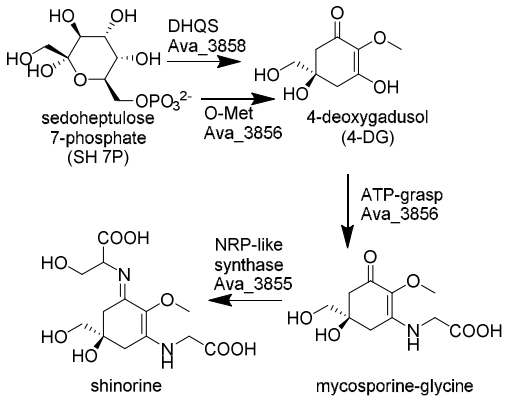


Figure 13: Mycosporine biosynthetic pathway. Dehydroquinate synthase (DHQS), O-methyltransferase (O-Met), nonribosomal peptide (NRP)-like synthase.

classes of natural products. The sulfotransferase catalytic cascade (ACP-ST-TE) of CurM from *Moorea producens* 3L and of the OLS pathway in *Synechococcus* PCC7002 was studied in detail by Gehret et al.[61, 64] The CurM ST and OLS ST have a low sequence similarity to previously characterized STs (less than 25% identity), underscoring the uniqueness of the cyanobacterial enzymes.[64] To test their function, substrates of different β-hydroxyacyl ACPs were incubated with the CurM ST and the OLS ST. Both STs preferred (3*R*)-3-hydroxyacyl-ACPs.[64] Interestingly, the OLS ST had no detectable activity towards the curacin A PKS ST substrate, which bears a 5-methoxy substituent. The crystal structures of both activating STs showed that they are similar to other STs in their overall architecture,[61] but show greatest similarity to one another. The key residues Glu60 and Lys133 form the anchoring site for the PAPS sulfate, and are invariant among activating STs.[61] The PAPS sulfate is thus located in an optimal location for nucleophilic attack by the β-hydroxy group of the substrate, which is deprotonated by Glu60.[61] After the activation of the β-hydroxy group, the TE facilitates terminal decarboxylation to form the alkene.[61] These STs are of great interest for natural product diversification and the creation of new bioenergetics materials. In addition to pathways that form hydrocarbons with terminal alkenes, cyanobacteria possess the capability to produce saturated alkanes using acyl-ACP reductases and aldehyde decarbonylases that operate on fatty-acyl-ACP substrates.[66, 67]

3.7 UV-induced natural products: Scytonemin and mycosporine-like amino acids

Scytonemin and mycosporine-like compounds are used by cyanobacteria for their UV-sunscreen properties, and their formation is trigged through exposure to UV light.[68, 69] It's clear that these compounds could have utility as sunscreen agents, especially given the increase UV irradiance reaching the planet surface as a result of climate change phenomena. Scytonemin is a dimeric indole phenolic pigment and is found in the sheaths of many cyanobacteria. Through isotopically labeled precursor feeding experiments it was shown that tyrosine and tryptophan are precursors of scytonemin biosynthesis.[70,71] The indolic precursor derives from tryptophan via oxidative deamination (NP1275, ScyB) to form indole-3-pyruvic acid (IPA). ScyA subsequently couples IPA to hydroxyphenylpyruvate, which is synthesized from prephenate using prephenate dehydrogenase (Np1269). The resulting β-ketoacid is cyclized by ScyC to form a tricyclic ketone resembling one-half of the scytonemin backbone. Through oxidative

transformations, including a dimerization reaction, scytonemin is produced.[68, 70, 72]

The mycosporines are a diverse family of water soluble, colorless compounds that have a 5-hydroxy-5-hydroxymethyl-cyclohex-1,2-ene ring and a C-2 methoxy substituent. At C-4 they are always substituted with an amino group and at C-1 they typically possess an oxo or imino group. In addition to cyanobacteria, these UV-absorbing pigments are produced by fungi and micro- and macro algae and even vertebrates.[73, 74] Their biosynthesis involves either precursors of the shikimate pathway or the pentose phosphate pathway, as reported from *in vitro* studies with *Anabaena variabilis* ATCC 29413 and *Nostoc punctiforme ATCC 29133*.[75, 76] Based on the pentose phosphate pathway, a 2-*epi*-5-*epi*-valiolone synthase (Ava_3585) catalyzes the transformation of sedoheptulose-7-phosphate to demethyl-4-deoxygadusol; this is subsequently *O*-methylated by an *O*-methyltransferase (Ava_3856) forming 4-deoxygadusol (4-DG). An ATP-grasp amino acid ligase catalyzes the addition of glycine to 4-DG forming mycosporine-glycine, and in turn, an NRPS-like synthetase catalyzes the ensuing condensation with serine to form shinorine.

3.8 Hapalindole-type natural products

Hapalindole-type compounds are terpenoid indole alkaloids that are produced by members of the Subsection V of cyanobacteria. To date, genome-sequencing efforts have detected the biosynthetic pathway genes in *Westiella intricate* UH strain HT-29-1, *Hapalosiphon welwitschii* UH strain IC-52-3 and UTEX B1830, *Fischerella ambigua* UTEX 1903, *Fischerella sp.* ATCC 43239 and PCC 9339 and *Fischerella musciola* SAG 1427-1.[77-79] The different types of hapalindol-type compounds can be summarized as hapalindoles[80], fischerindoles[81], welwitindolinones[82] and ambiguines (Figure 14).[79] A total of 19 core genes have been identified; the tryptophan biosynthesis genes *T1-5* and *C2*, the isonitrile biosynthesis genes *I1-3*, the isoprenoid biosynthesis genes *D1-4*, the geranyl pyrophosphate synthase gene *P2*, the hapalindole-specific aromatic prenyltransferase gene *P1*, the regulatory protein-encoding genes *R1* and *R2*, as well as *C1* and *C3* which encode for other enzymes. In the cyanobacterial species analyzed in this study, a greater than 92% sequence identity at the protein level was observed when comparing each of these 19 genes with their respective homologs.[77] Additional oxygenase genes, such as the Rieske oxygenase, are present but their biochemical role is unknown. Hillwig et al. performed *in vitro* studies to characterize key enzymes in the biosynthesis of ambiguine and welwitindolinoe, and clearly linked this pathway to production of hapalindole-type natural products.[78, 79] They also studied the role of WelO5, a nonheme iron enzyme present in the welwitindolinone biosynthetic pathway, and demonstrated its capacity to monochlorinate an aliphatic carbon in 12-*epi*-fischerindole U and 12-*epi*-hapalindole C.[58] The logic implied by the organization of biosynthetic gene clusters for these metabolites does not follow the co-linearity rule so typical of PKS and NRPS metabolites.

3.9 Prenylation and macrocyclization of ribosomal peptides

Cyanobacteria also produce ribosomal peptides, which are then subject to several interesting post-translational modifications.[83] Prominent examples are the cyanobactins.[84, 85] Intriguingly, these ribosomal peptides are often found to be *C*-prenylated or *O*-prenylated, but rarely N-prenylated.[85-87] A recent study examined the prenyltransferase LynF (aesturamide pathway), and provided fundamental insights into the forward *C*-prenylation mechanism.[88, 89] LynF was isolated from *Lyngbya aestuarii* and belongs to the TruF family. The TruF prenyltransferase family is responsible for both reverse- and forward-*O*-prenylation of tyrosine, serine and threonine residues in cyclic peptides using dimethylallyl pyrophosphate (DMAPP) as the prenylation precursor (Figure 15).[89] These prenyltransferases possess a broad spectrum of substrate specificity, making them of interest for biochemical engineering.[89]

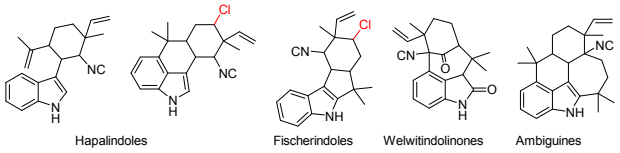


Figure 14: Hapalindole-type natural products

As demonstrated *in vitro* by McIntosh et al.,[88, 89] *C*-prenylation of phenols is not enzymatic, but rather, is based on a Claisen rearrangement from the *O*-prenylated precursor.

The interaction of RiPP precursor peptides with maturation enzymes was recently elucidated by the crystal structure of one forming heterocycles as well as another forming a macro-cycle.[90-92] The ATP-dependent enzyme LynD (aesturamide pathway) heterocyclizes multiple cysteine residues to thiazolines within a peptide substrate and requires the substrate to have a conserved N-terminal leader sequence for full activity.[91] Next, a LynD variant was engineered which heterocyclizes substrates without a leader peptide present.[91] In another study, the mechanism of macrocyclization of the patellamide pathway (Pat) was investigated by overexpressing PatGmac (residues 492-851) yielding cycloVGAGIGFP.[92] PatGmac has a catalytic triad of Asp548, His618 and Ser783 which demonstrates that PatGmac belongs to the subtilisin class of proteases.[92] Overall, the capability to macrocyclize a linear peptide precursor has great potential for biotechnology applications.

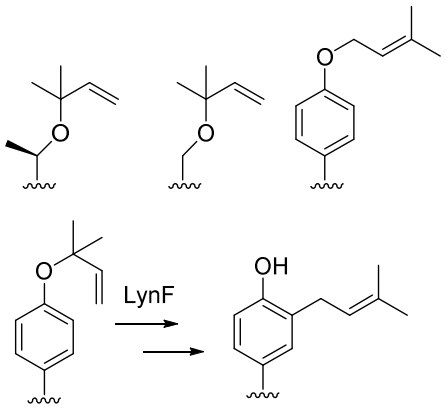


Figure 15: Ribosomal peptide prenylation.

4. Heterologous expression of cyanobacterial gene clusters and chemo-enzymatic synthesis of the cryptophycins

It is a considerable challenge to perform genetic manipulations in filamentous cyanobacteria because they grow slowly under laboratory conditions (doubling rates of 6 days or more) and are encased in a thick polysaccharide sheath that inhibits the introduction of DNA by conventional laboratory transformation or conjugation methods. The sheath also harbors a rich assortment of associated heterotrophic bacteria.[31, 37, 93] Therefore, the development of heterologous expression systems for these organisms is much needed. In two recent cases, cyanobacterial natural product PKS/NRPS biosynthetic pathways were heterologously expressed and led to the production of 4-*O*-demethylbarbamide and lyngbyatoxin, respectively. In addition, chemoenzymatic reactions of the cryptophycin pathway make it possible to create new analogs with improved biological activity.

4.1 Heterologous expression of the barbamide biosynthetic pathway

The barbamide biosynthetic pathway is a hybrid NRPS/PKS pathway with unique tailoring reactions such as trichlorination of an unactivated methyl group, a one-carbon truncation during chain extension, formation of an *E*-double bound, and thiazole ring formation.[94] The 26 kb gene cluster contains 12 ORFs, *barA-barK*.[95] These were cloned into a replicative *E. coli – Streptomyces* shuttle vector pDHS702, containing the *pikAI* promotor from the pikromycin PKS pathway. The *barA* - *barH* genes were under transcriptional control by the *pik*AI promotor whereas *barI-barK* were under control by the native barbamide promotor. The native intergenic regions and ribosome-binding site of each gene were preserved. The engineered plasmid was cloned into *S. venezuelae* DHS 2001, a promising host for the heterologous expression of polyketides, hybrid polyketide-nonribosomal peptides and aminoglycosides due to its rapid growth and relative ease of genetic manipulation.[94] The new strain *S. venezuelae* YJ348 produced a

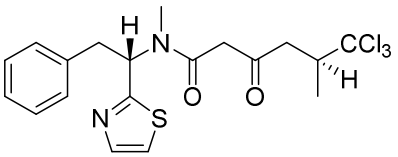
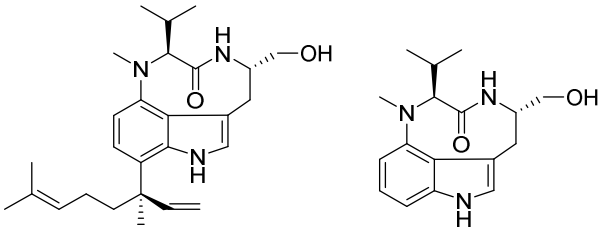


Figure 16: 4-O-Demethylbarbamide

barbamide-related compound, 4-*O*-demethylbarbamide, which was subsequently found to be a natural product in cultures of *M. producens* 3L. For unknown reasons, the BarF-catalyzed *O*-methylation reaction does not function in this heterologous expression system.[94] It is interesting to note that the native promotor of *Moorea* was recognized by *S. venezuelae* since transcripts of *barJ* were detected. Unfortunately, the yield was very low, less than 1 µg/L, which may be due to suboptimal codon usage in cyanobacterial DNA compared to the high GC content of actinobacteria. Nevertheless, heterologous expression of the barbamide pathway represents an important milestone in efforts to heterologously express cyanobacterial biosynthetic pathways.[94]

4.2 Heterologous expression of lyngbyatoxin A

In 2013 high titers of lyngbyatoxin A (25.6 mg/L) and its precursor indolactam-V (150 mg/L) were produced by an *Escherichia coli* strain into which a plasmid containing the lyngbyatoxin gene cluster



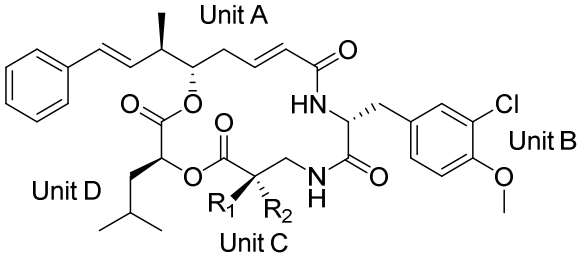
Lyngbyatoxin A Indolactam-V

Figure 17: Lyngbyatoxin A and Indolactam-V

had been inserted. The high expression levels may be in part due to the relatively similar %GC content between the *E. coli* strain and the lyngbyatoxin biosynthetic pathway (AY588942).[96] Lyngbyatoxin A is an indole alkaloid that promotes tumors by activating protein kinase C. It was isolated from a cyanobacterial bloom off the Hawaiian island of Oahu where it occasionally causes a skin irritation in bathers known as 'swimmer's itch'. The 11.3 kb biosynthetic pathway has four open reading frames (*ltxA-D*). The first, LtxA, is a bimodular NRPS that condenses L-methyl-valine and L-tryptophan to form the linear dipeptide *N*-methyl-L-valyl-L-tryptophan. This is released as a primary alcohol through a NADPH-dependent reductive cleavage. LtxB, a P450-dependent monooxygenase catalyzes the formation of indolactam-V through the oxidation and cyclization of the dipeptide.[97] Finally, the reverse prenyltransferase LtxC uses geranyl pyrophosphate to add the terpene moiety to the C-7 position. It is proposed that lyngbyatoxin B and C form via oxidation catalyzed by the oxidoreductase LtxD. Successful expression of lyngbyatoxin was accomplished in *E. coli* GB05-MtaA,[98, 99] a strain that has been transfected with a myxobacterial PPTase of broad-range substrate activity. This strain has found utility in the expression of NRPS clusters that are under control by the tetO tetracycline-inducible promotor (P_{tetO}). With such a promoter installed to drive the ltx gene construct, and at the reduced temperature of 18 °C, significant functional expression of the pathway was observed.

4.3 Chemoenzymatic synthesis of cryptophycin analogues

Advances in heterologous expression of cyanobacterial compounds, including the re-engineering of problematic biosynthetic reactions, will help improve product yields and make bioactive analogs more easily available. For example, the cryptophycins from *Nostoc* sp.[100] are cyanobacterial depsipeptides with promising activity against drug-resistant tumors. To optimize their medicinal and pharmaceutical properties, the production of analogs was undertaken using synthetic chemistry methods.[101] The cryptophycin biosynthetic gene cluster was characterized to possess two type I PKS genes, *crpA* and *crpB*, two nonribosomal peptide synthetase (NRPS) genes, *crpC* and *crpD* and four tailoring enzyme genes,



Cryptophycin 3: R1=Me, R2=H
Cryptophycin 24: R1=H, R2=H
Cryptophycin 51: R1=Me, R2=Me

Figure 18: Cryptophycin

including a P450 epoxidase, *crpE*. Chemoenzymatic efforts with this pathway were used to make a number of analogs, and involved macrocyclization[102], regio- and stereo-specifically epoxidation[103] and formation of ester bonds using one of the nonribosomal peptide synthetase enzymes.[104] Recently, chemoenzymatic synthesis of cryptophycin 3, 24 and 51 using CrpD-M2 and the Crp thioesterase worked in a proof-of-principle effort to create new cryptophycin analogues with unnatural elements in the C and D structural units (Figure 17).[104] CrpD-M2 is an ester bond-forming nonribosomal peptide synthetase that incorporates a 2-hydroxy acid moiety at unit D (Figure 18). It consists of condensation, adenylation, ketoreduction and thiolation (C-A-KR-T) domains. The CrpD-M2 adenylation domain has a similar selectivity toward 2-hydroxy and 2-keto-acids as shown by ATP-PPi exchange assays that demonstrated substrate specificity for 2-keto-isocaproate, 2-keto-γ-(methylthio)-butyrate, L-2-hydroxy-isocaproate, 2-oxovalearate and 3-methyl-2-oxo-valerate.[104] Only $HctE_{IVA}$ from the hectochlorin pathway has a similar selectivity for 2-keto and 2-hydroxy acids.[33] The KR domain of CrpD-M2 reduces 2-keto-isocaproate to L-2-hydroxy-isocaporate after the addition of reducing cofactors (e.g. NADPH, NADH). CrpD-M2 was successfully used to form *seco*-cryptophycin intermediates as well as cyclized cryptophycins 3, 24 and 51 using synthetic SNAC-ABC chain elongation intermediates as substrates.[105, 106] This was the first study in which a NRPS with an embedded KR domain was utilized to generate bioactive compounds from natural and synthetic chain elongation intermediates, and should enable production of new cryptophycin analogous with improved bioactivity.

5. Phylum-wide genome sequencing projects

Phylum-wide genome sequencing together with enhanced bioinformatic tools will help to identify new natural product biosynthetic capabilities as well as the breadth of natural product-producing strains. Additionally, such efforts will help characterize and appreciate the roles of microbial diversity in cyanobacterial assemblages, and also to distinguish between the many cyanobacterial chemotypes.[107-109] The recent sequencing effort known as the "CyanoGEBA project" sought to increase the diversity of cyanobacterial genomes by sequencing 126 species, and resulted in the identification of more than 21,000 cyanobacterial proteins with no detectable similarity to previously known proteins.[110] Follow-up studies on secondary metabolite capacity within 89 of these newly sequence genomes revealed 452 NRPS and PKS gene clusters. These grouped into 286 highly diversified biosynthetic gene cluster families.[13] The products of only 20% of these are known compounds, underscoring the potential to find new and unique natural products in these organisms.[13] Detection of shared pathways at the phylum level provides a valuable tool for prioritization of future compound identification projects and sheds light on the functional expression of these biosynthetic pathways.[111]

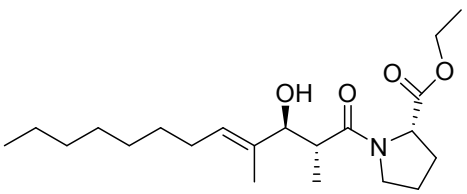


Figure 19: Ethyl tumonoate

An exceptionally secondary metabolite rich group of cyanobacteria is the tropical genus *Moorea*, formerly known as *Lyngbya.*[31, 37, 112] Somewhat greater than 40% of all reported marine cyanobacterial natural products have been reported from this genus,[112] and genomes of several strains have been sequenced and will be published shortly. However, due to the presence of heterotrophic bacteria associated with their sheaths and the fact that these organisms exist in Nature as complex cyanobacterial assemblages, combined with the lack of effective heterologous expression systems, it is still quite difficult to obtain complete genome sequences as well as to functionally characterize these pathways. Nevertheless, advances are being made and utilized to help in the discovery of new and biologically active secondary metabolites. For example, a phylogeny-guided isolation of cyanobacterial strains from Curaçao enabled the discovery of ethyl tumonoate A, a new tumonoic acid derivative with anti-inflammatory activity in murine macrophage cells (Figure 19).[107] In this case, a direct correlation between phylogenetics of *Oscillatoria margaritifera* strains and secondary metabolite production was observed.[107]

6. Conclusion

Many marine life forms have emerged as exceptionally rich sources of secondary metabolites with intriguing biological and chemical properties. The combination of genome sequencing together with the top down bioactivity-guided isolation and innovative mass spectrometric dereplication techniques is helping to facilitate the drug discovery process. Unique gene "tags", such as the β-branching cassettes discovered in the curacin A, jamaicamide and apratoxin pathways, or the terminal alkyne producing enzymes present in the carmabin and jamaicamide pathways, are helping to accelerate the search of recently sequenced cyanobacterial genomes for novel chemical entities. Understanding the biochemical nature of novel enzymatic transformations will certainly assist in the development of new bioengineered compounds. An important advance in this field has been to develop an understanding of the fidelity of inter-module docking domains which allows for the efficient transfer of substrates between modules, a key aspect of effective heterologous expression and pathway engineering. Similarly, combining the effectiveness of atom assembly accomplished by synthetic organic chemistry with the regio- and stereo-control of enzymes, as demonstrated in the cryptophycin case, represents a new milestone in natural product bioengineering. In the course of investigations of these biosynthetic systems, Nature has demonstrated a wide diversity of approaches to natural product pathway diversification and resulting production of bioactive metabolites; now it is our turn to use this knowledge to create new compounds with useful and exciting biological properties.

7. References

1. W. H. Gerwick and B. S. Moore, *Chem. Biol.*, 2011, **19**, 85-98.
2. B. Rasmussen, I. R. Fletcher, J. J. Brocks and M. R. Kilburn, *Nature*, 2008, **455**, 1101-1104.
3. H. W. Paerl, J. L. Pinckney and T. F. Steppe, *Environ. Microbiol.*, 2000, **2**, 11-26.
4. L. A. Salvador-Reyes and H. Luesch, *Nat. Prod. Rep.*, 2015, **32**, 478-503.
5. R. B. Dixit and M. R. Suseela, *Anton. Leeuw.*, 2013, **103**, 947-961.
6. E. D. Hilborn and V. R. Beasley, *Toxins*, 2015, **7**, 1374-1395.

7. J. W. Blunt, B. R. Copp, W. P. Hu, M. H. G. Munro, P. T. Northcote and M. R. Prinsep, *Nat. Prod. Rep.*, 2009, **26**, 170-244.
8. J. A. Kalaitzis, F. M. Lauro and B. A. Neilan, *Nat. Prod. Rep.*, 2009, **26**, 1447-1465.
9. W. H. Gerwick, P. J. Proteau, D. G. Nagle, E. Hamel, A. Blokhin and D. L. Slate, *J. Org. Chem.*, 1994, **59**, 1243-1245.
10. P. Verdier-Pinard, J. Y. Lai, H. D. Yoo, J. Yu, B. Marquez, D. G. Nagle, M. Nambu, J. D. White, J. R. Falck, W. H. Gerwick, B. W. Day and E. Hamel, *Mol. Pharmacol.*, 1998, **53**, 62-76.
11. L. Gu, B. Wang, A. Kulkarni, T. W. Geders, R. V. Grindberg, L. Gerwick, K. Hakansson, P. Wipf, J. L. Smith, W. H. Gerwick and D. H. Sherman, *Nature*, 2009, **459**, 731-735.
12. R. V. Grindberg, T. Ishoey, D. Brinza, E. Esquenazi, R. C. Coates, W. T. Liu, L. Gerwick, P. C. Dorrestein, P. Pevzner, R. Lasken and W. H. Gerwick, *PLoS One*, 2011, **6**, e18565.
13. A. Calteau, D. P. Fewer, A. Latifi, T. Coursin, T. Laurent, J. Jokela, C. A. Kerfeld, K. Sivonen, J. Piel and M. Gugger, *BMC Genomics*, 2014, **15**, 977.
14. A. C. Jones, E. A. Monroe, E. B. Eisman, L. Gerwick, D. H. Sherman and W. H. Gerwick, *Nat. Prod. Rep.*, 2010, **27**, 1048-1065.
15. N. Engene, S. P. Gunasekera, W. H. Gerwick and V. J. Paul, *Appl. Environ. Microbiol.*, 2013, **79**, 1882–1888.
16. D. J. Edwards, B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Goeger, M. A. Roberts and W. H. Gerwick, *Chem. Biol.*, 2004, **11**, 817-833.
17. H. Luesch, W. Y. Yoshida, R. E. Moore, V. J. Paul and T. H. Corbett, *J. Am. Chem. Soc.*, 2001, **123**, 5418-5423.
18. H. Luesch, S. K. Chanda, R. M. Raya, P. D. DeJesus, A. P. Orth, J. R. Walker, J. C. Izpisua Belmonte and P. G. Schultz, *Nat. Chem. Biol.*, 2006, **2**, 158-167.
19. Z. Chang, N. Sitachitta, J. V. Rossi, M. A. Roberts, P. M. Flatt, J. Jia, D. H. Sherman and W. H. Gerwick, *J. Nat. Prod.*, 2004, **67**, 1356-1367.
20. L. Gu, J. Jia, H. Liu, K. Hakansson, W. H. Gerwick and D. H. Sherman, *J. Am. Chem. Soc.*, 2006, **128**, 9014-9015.
21. T. W. Geders, L. Gu, J. C. Mowers, H. Liu, W. H. Gerwick, K. Hakansson, D. H. Sherman and J. L. Smith, *J. Biol. Chem.*, 2007, **282**, 35954-35963.
22. K. Kleigrewe, J. Almaliti, I. Y. Tian, R. B. Kinnel, A. Korobeynikov, E. A. Monroe, B. M. Duggan, V. Di Marzo, D. H. Sherman, P. C. Dorrestein, L. Gerwick and W. H. Gerwick, *J. Nat. Pro.*, 2015, **78**, 1671-1682.
23. J. Watrous, P. Roach, T. Alexandrov, B. S. Heath, J. Y. Yang, R. D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J. M. Raaijmakers, B. S. Moore, J. Laskin, N. Bandeira and P. C. Dorrestein, *Proc. Natl. Acad. Sci. USA*, 2012, **109**, E1743-1752.
24. J. R. Winnikoff, E. Glukhov, J. Watrous, P. C. Dorrestein and W. H. Gerwick, *J. Antibiot.*, 2014, **67**, 105-112.
25. H. Nakamura and E. P. Balskus, *Synlett*, 2013, **24**, 1464-1470.
26. H. Nakamura, H. A. Hamer, G. Sirasani and E. P. Balskus, *J. Am. Chem. Soc.*, 2012, **134**, 18518-18521.
27. P. N. Leao, H. Nakamura, M. Costa, A. R. Pereira, R. Martins, V. Vasconcelos, W. H. Gerwick and E. P. Balskus, *Angew. Chem. Int. Ed. Engl.*, 2015, **54**, 11063-11067.
28. D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick and J. L. Smith, *Structure*, 2010, **18**, 94-105.
29. D. J. Edwards and W. H. Gerwick, *J. Am. Chem. Soc.*, 2004, **126**, 11432-11433.
30. B. L. Marquez, K. S. Watts, A. Yokochi, M. A. Roberts, P. Verdier-Pinard, J. I. Jimenez, E. Hamel, P. J. Scheuer and W. H. Gerwick, *J. Nat. Prod.*, 2002, **65**, 866-871.
31. N. Engene, E. C. Rottacker, J. Kastovsky, T. Byrum, H. Choi, M. H. Ellisman, J. Komarek and W. H. Gerwick, *Int. J. Syst. Evol. Microbiol.*, 2012, **62**, 1171-1178.
32. E. A. Monroe, H. Choi, V. Lesin, A. Sirotkin, M. Dvorkin, P. Pevzner, W. H. Gerwick and L. Gerwick, *Planta Med.*, 2012, **78**, CL52.
33. A. V. Ramaswamy, C. M. Sorrels and W. H. Gerwick, *J. Nat. Prod.*, 2007, **70**, 1977-1986.
34. H. Luesch, W. Y. Yoshida, R. E. Moore, V. J. Paul and S. L. Mooberry, *J. Nat. Prod.*, 2000, **63**, 611-615.
35. K. L. McPhail, J. Correa, R. G. Linington, J. Gonzalez, E. Ortega-Barria, T. L. Capson and W. H. Gerwick, *J. Nat. Prod.*, 2007, **70**, 984-988.
36. G. J. Hooper, J. Orjala, R. C. Schatzman and W. H. Gerwick, *J. Nat. Prod.*, 1998, **61**, 529-533.
37. A. C. Jones, E. A. Monroe, S. Podell, W. R. Hess, S. Klages, E. Esquenazi, S. Niessen, H. Hoover, M. Rothmann, R. S. Lasken, J. R. Yates, 3rd, R. Reinhardt, M. Kube, M. D. Burkart, E. E. Allen, P. C. Dorrestein, W. H. Gerwick and L. Gerwick, *Proc. Natl. Acad. Sci. USA*, 2011, **108**, 8815-8820.
38. X. Zhu, J. Liu and W. Zhang, *Nat. Chem. Biol.*, 2015, **11**, 115-120.
39. E. Dittmann, M. Gugger, K. Sivonen and D. P. Fewer, *Trends microbiol.*, 2015, **23**, 642-652.
40. J. W. Blunt, B. R. Copp, R. A. Keyzers, M. H. Munro and M. R. Prinsep, *Nat. Prod. Rep.*, 2015, **32**, 116-211.
41. D. L. Akey, J. J. Gehret, D. Khare and J. L. Smith, *Nat. Prod. Rep.*, 2012, **29**, 1038-1049.
42. L. Gu, E. B. Eisman, S. Dutta, T. M. Franzmann, S. Walter, W. H. Gerwick, G. Skiniotis and D. H. Sherman, *Angew. Chem. Int. Ed. Engl.*, 2011, **50**, 2795-2798.
43. J. Crosby and M. P. Crump, *Nat. Prod. Rep.*, 2012, **29**, 1111-1137.
44. R. A. Butcher, F. C. Schroeder, M. A. Fischbach, P. D. Straight, R. Kolter, C. T. Walsh and J. Clardy, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 1506-1509.
45. A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson and C. M. Thomas, *Chem. Biol.*, 2003, **10**, 419-430.
46. J. Piel, D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani and S. Matsunaga, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 16222-16227.
47. H. Jiang, R. Zirkle, J. G. Metz, L. Braun, L. Richter, S. G. Van Lanen and B. Shen, *J. Am. Chem. Soc.*, 2008, **130**, 6336-6337.
48. A. S. Rahman, J. Hothersall, J. Crosby, T. J. Simpson and C. M. Thomas, *J. Biol. Chem.*, 2005, **280**, 6399-6408.
49. A. S. Haines, X. Dong, Z. Song, R. Farmer, C. Williams, J. Hothersall, E. Ploskon, P. Wattana-amorn, E. R. Stephens, E. Yamada, R. Gurney, Y. Takebayashi, J. Masschelein, R. J. Cox, R. Lavigne, C. L. Willis, T. J. Simpson, J. Crosby, P. J. Winn, C. M. Thomas and M. P. Crump, *Nat. Chem. Biol.*, 2013, **9**, 685-692.
50. A. Busche, D. Gottstein, C. Hein, N. Ripin, I. Pader, P. Tufar, E. B. Eisman, L. Gu, C. T. Walsh, D. H. Sherman, F. Löhr, P. Güntert and V. Dötsch, *ACS Chem. Biol.*, 2012, **7**, 378-386.
51. T. J. Buchholz, T. W. Geders, F. E. Bartley, K. A. Reynolds, J. L. Smith and D. H. Sherman, *ACS Chem. Biol.*, 2009, **4**, 41-52.
52. J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W. H. Gerwick, D. H. Sherman and J. L. Smith, *Chem. Biol.*, 2013, **20**, 1340-1351.
53. A. Keatinge-Clay, *J. Mol. Biol.*, 2008, **384**, 941-953.
54. R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson and R. McDaniel, *Biochemistry*, 2003, **42**, 72-79.

55. D. Khare, B. Wang, L. Gu, J. Razelun, D. H. Sherman, W. H. Gerwick, K. Hakansson and J. L. Smith, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 14099-14104.
56. D. P. Galonic, E. W. Barr, C. T. Walsh, J. M. Bollinger, Jr. and C. Krebs, *Nat. Chem. Biol.*, 2007, **3**, 113-116.
57. S. M. Pratter, J. Ivkovic, R. Birner-Gruenberger, R. Breinbauer, K. Zangger and G. D. Straganz, *ChemBioChem*, 2014, **15**, 567-574.
58. M. L. Hillwig and X. Y. Liu, *Nat. Chem. Biol.*, 2014, **10**, 921-923.
59. S. M. Pratter, K. M. Light, E. I. Solomon and G. D. Straganz, *J. Am. Chem. Soc.*, 2014, **136**, 9385-9395.
60. J. J. Gehret, L. Gu, T. W. Geders, W. C. Brown, L. Gerwick, W. H. Gerwick, D. H. Sherman and J. L. Smith, *Protein Sci.*, 2012, **21**, 239-248.
61. J. G. McCarthy, E. B. Eisman, S. Kulkarni, L. Gerwick, W. H. Gerwick, P. Wipf, D. H. Sherman and J. L. Smith, *ACS Chem. Biol.*, 2012, **7**, 1994-2003.
62. D. Mendez-Perez, M. B. Begemann and B. F. Pfleger, *Appl. Environ. Microbiol.*, 2011, **77**, 4264-4267.
63. R. C. Coates, S. Podell, A. Korobeynikov, A. Lapidus, P. Pevzner, D. H. Sherman, E. E. Allen, L. Gerwick and W. H. Gerwick, *PLoS One*, 2014, **9**, e85140.
64. J. J. Gehret, L. Gu, W. H. Gerwick, P. Wipf, D. H. Sherman and J. L. Smith, *J. Biol. Chem.*, 2011, **286**, 14445-14454.
65. L. Gu, B. Wang, A. Kulkarni, J. J. Gehret, K. R. Lloyd, L. Gerwick, W. H. Gerwick, P. Wipf, K. Hakansson, J. L. Smith and D. H. Sherman, *J. Am. Chem. Soc.*, 2009, **131**, 16033-16035.
66. T. Yoshino, Y. Liang, D. Arai, Y. Maeda, T. Honda, M. Muto, N. Kakunaka and T. Tanaka, *Appl. Microbiol. Biotechnol.*, 2015, **99**, 1521-1529.
67. A. Schirmer, M. A. Rude, X. Li, E. Popova and S. B. del Cardayre, *Science*, 2010, **329**, 559-562.
68. E. P. Balskus, R. J. Case and C. T. Walsh, *FEMS Microbiol. Ecol.*, 2011, **77**, 322-332.
69. Q. Gao and F. Garcia-Pichel, *Nat. Rev. Microbiol.*, 2011, **9**, 791-802.
70. T. Soule, K. Palmer, Q. J. Gao, R. M. Potrafka, V. Stout and F. Garcia-Pichel, *Bmc Genomics*, 2009, **10**.
71. C. S. Jones, E. Esquenazi, P. C. Dorrestein and W. H. Gerwick, *Bioorg. Med. Chem.*, 2011, **19**, 6620-6627.
72. C. M. Sorrels, P. J. Proteau and W. H. Gerwick, *Appl. Environ. Microbiol.*, 2009, **75**, 4861-4869.
73. W. M. Bandaranayake, *Nat. Prod. Rep.*, 1998, **15**, 159-172.
74. A. R. Osborn, K. H. Almabruk, G. Holzwarth, S. Asamizu, J. LaDu, K. M. Kean, P. A. Karplus, R. L. Tanguay, A. T. Bakalinsky and T. Mahmud, *Elife*, 2015, **12**, 05919.
75. M. A. Pope, E. Spence, V. Seralvo, R. Gacesa, S. Heidelberger, A. J. Weston, W. C. Dunlap, J. M. Shick and P. F. Long, *Chembiochem*, 2015, **16**, 320-327.
76. E. P. Balskus and C. T. Walsh, *Science*, 2010, **329**, 1653-1656.
77. M. L. Micallef, D. Sharma, B. M. Bunn, L. Gerwick, R. Viswanathan and M. C. Moffitt, *BMC Microbiol.*, 2014, **14**, 014-0213.
78. M. L. Hillwig, H. A. Fuhrman, K. Ittiamornkul, T. J. Sevco, D. H. Kwak and X. Liu, *Chembiochem*, 2014, **15**, 665-669.
79. M. L. Hillwig, Q. Zhu and X. Liu, *ACS Chem. Biol.*, 2013, **9**, 372-377.
80. R. E. Moore, C. Cheuk and G. M. L. Patterson, *J. Am. Chem. Soc.*, 1984, **106**, 6456-6457.
81. A. Park, R. E. Moore and G. M. L. Patterson, *Tetrahedron Lett.*, 1992, **33**, 3257-3260.
82. K. Stratmann, R. E. Moore, R. Bonjouklian, J. B. Deeter, G. M. L. Patterson, S. Shaffer, C. D. Smith and T. A. Smitka, *J. Am. Chem. Soc.*, 1994, **116**, 9935-9942.
83. P. G. Arnison, M. J. Bibb, G. Bierbaum, A. A. Bowers, T. S. Bugni, G. Bulaj, J. A. Camarero, D. J. Campopiano, G. L. Challis, J. Clardy, P. D. Cotter, D. J. Craik, M. Dawson, E. Dittmann, S. Donadio, P. C. Dorrestein, K. D. Entian, M. A. Fischbach, J. S. Garavelli, U. Goransson, C. W. Gruber, D. H. Haft, T. K. Hemscheidt, C. Hertweck, C. Hill, A. R. Horswill, M. Jaspars, W. L. Kelly, J. P. Klinman, O. P. Kuipers, A. J. Link, W. Liu, M. A. Marahiel, D. A. Mitchell, G. N. Moll, B. S. Moore, R. Muller, S. K. Nair, I. F. Nes, G. E. Norris, B. M. Olivera, H. Onaka, M. L. Patchett, J. Piel, M. J. Reaney, S. Rebuffat, R. P. Ross, H. G. Sahl, E. W. Schmidt, M. E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein, R. D. Sussmuth, J. R. Tagg, G. L. Tang, A. W. Truman, J. C. Vederas, C. T. Walsh, J. D. Walton, S. C. Wenzel, J. M. Willey and W. A. van der Donk, *Nat. Prod. Rep.*, 2013, **30**, 108-160.
84. J. A. McIntosh, M. S. Donia and E. W. Schmidt, *Nat. Prod. Rep.*, 2009, **26**, 537-559.
85. Mohamed S. Donia and Eric W. Schmidt, *Chem. Biol.*, 2011, **18**, 508-519.
86. M. S. Donia, J. Ravel and E. W. Schmidt, *Nat. Chem. Biol.*, 2008, **4**, 341-343.
87. N. Leikoski, L. Liu, J. Jokela, M. Wahlsten, M. Gugger, A. Calteau, P. Permi, C. A. Kerfeld, K. Sivonen and D. P. Fewer, *Chem. Biol.*, 2013, **20**, 1033-1043.
88. J. A. McIntosh, Z. Lin, M. D. B. Tianero and E. W. Schmidt, *ACS Chem. Biol.*, 2013, **8**, 877-883.
89. J. A. McIntosh, M. S. Donia, S. K. Nair and E. W. Schmidt, *J. Am. Chem. Soc.*, 2011, **133**, 13698-13705.
90. A. J. Link, *Nature Chemical Biology*, 2015, **11**, 551-552.
91. J. Koehnke, G. Mann, A. F. Bent, H. Ludewig, S. Shirran, C. Botting, T. Lebl, W. E. Houssen, M. Jaspars and J. H. Naismith, *Nat. Chem. Biol.*, 2015, **11**, 558-563.
92. J. Koehnke, A. Bent, W. E. Houssen, D. Zollman, F. Morawitz, S. Shirran, J. Vendome, A. F. Nneoyiegbe, L. Trembleau, C. H. Botting, M. C. Smith, M. Jaspars and J. H. Naismith, *Nat. Struct. Mol. Biol.*, 2012, **19**, 767-772.
93. E. Esquenazi, A. C. Jones, T. Byrum, P. C. Dorrestein and W. H. Gerwick, *Proc. Natl. Acad. Sci. USA*, 2011, **108**, 5226-5231.
94. E. J. Kim, J. H. Lee, H. Choi, A. R. Pereira, Y. H. Ban, Y. J. Yoo, E. Kim, J. W. Park, D. H. Sherman, W. H. Gerwick and Y. J. Yoon, *Org. Lett.*, 2012, **14**, 5824-5827.
95. Z. Chang, P. Flatt, W. H. Gerwick, V. A. Nguyen, C. L. Willis and D. H. Sherman, *Gene*, 2002, **296**, 235-247.
96. S. E. Ongley, X. Bian, Y. Zhang, R. Chau, W. H. Gerwick, R. Müller and B. A. Neilan, *ACS Chem. Biol.*, 2013.
97. M. U. Huynh, M. C. Elston, N. M. Hernandez, D. B. Ball, S. Kajiyama, K. Irie, W. H. Gerwick and D. J. Edwards, *J. Nat. Prod.*, 2010, **73**, 71-74.
98. X. Bian, F. Huang, F. A. Stewart, L. Xia, Y. Zhang and R. Müller, *Chembiochem*, 2012, **13**, 1946-1952.
99. J. Fu, X. Bian, S. Hu, H. Wang, F. Huang, P. M. Seibert, A. Plaza, L. Xia, R. Muller, A. F. Stewart and Y. Zhang, *Nat. Biotechnol.*, 2012, **30**, 440-446.
100. N. A. Magarvey, Z. Q. Beck, T. Golakoti, Y. Ding, U. Huber, T. K. Hemscheidt, D. Abelson, R. E. Moore and D. H. Sherman, *ACS Chem. Biol.*, 2006, **1**, 766-779.
101. C. Weiss, B. Sammet and N. Sewald, *Nat. Prod. Rep.*, 2013, **30**, 924-940.
102. Z. Q. Beck, C. C. Aldrich, N. A. Magarvey, G. I. Georg and D. H. Sherman, *Biochemistry*, 2005, **44**, 13457-13466.

103. Y. Ding, W. H. Seufert, Z. Q. Beck and D. H. Sherman, *J. Am. Chem. Soc.*, 2008, **130**, 5492-5498.
104. Y. Ding, C. M. Rath, K. L. Bolduc, K. Håkansson and D. H. Sherman, *J. Am. Chem. Soc.*, 2011, **133**, 14492-14495.
105. K. L. Bolduc, S. D. Larsen and D. H. Sherman, *Chem. Commun.*, 2012, **48**, 6414-6416.
106. W. Seufert, Z. Q. Beck and D. H. Sherman, *Angew. Chem. Int. Ed. Engl.*, 2007, **46**, 9298-9300.
107. N. Engene, H. Choi, E. Esquenazi, T. Byrum, F. A. Villa, Z. Cao, T. F. Murray, P. C. Dorrestein, L. Gerwick and W. H. Gerwick, *J. Nat. Prod.*, 2011, **74**, 1737-1743.
108. K. Blin, M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano and T. Weber, *Nucleic Acids Res.*, 2013, **41**, W204-W212.
109. E. Dittmann, D. P. Fewer and B. A. Neilan, *FEMS Microbiol. Rev.* , 2013, **37**, 23-43.
110. P. M. Shih, D. Wu, A. Latifi, S. D. Axen, D. P. Fewer, E. Talla, A. Calteau, F. Cai, N. Tandeau de Marsac, R. Rippka, M. Herdman, K. Sivonen, T. Coursin, T. Laurent, L. Goodwin, M. Nolan, K. W. Davenport, C. S. Han, E. M. Rubin, J. A. Eisen, T. Woyke, M. Gugger and C. A. Kerfeld, *Proc. Natl. Acad. Sci. U S A*, 2013, **110**, 1053-1058.
111. M. L. Micallef, P. M. D'Agostino, B. Al-Sinawi, B. A. Neilan and M. C. Moffitt, *Mar. Genomics*, 2014, **21**, 1-12.
112. N. Engene, H. Choi, E. Esquenazi, E. C. Rottacker, M. H. Ellisman, P. C. Dorrestein and W. H. Gerwick, *Environ. Microbiol.*, 2011, **13**, 1601-1610.