Published in final edited form as:

Nat Prod Rep. 2012 October ; 29(10): 1038–1049. doi:10.1039/c2np20016c.

Insights from the Sea: Structural Biology of Marine Polyketide Synthases

David L. Akey¹, Jennifer J. Gehret^{1,2}, Dheeraj Khare¹, and Janet L. Smith^{1,2,*}

¹Life Sciences Institute, University of Michigan, Ann Arbor, MI 48109

²Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109

Abstract

The world's oceans are a rich source of natural products with extremely interesting chemistry. Biosynthetic pathways have been worked out for a few, and the story is being enriched with crystal structures of interesting pathway enzymes. By far, the greatest number of structural insights from marine biosynthetic pathways has originated with studies of curacin A, a poster child for interesting marine chemistry with its cyclopropane and thiazoline rings, internal *cis* double bond, and terminal alkene. Using the curacin A pathway as a model, structural details are now available for a novel loading enzyme with remarkable dual decarboxylase and acetyltransferase activities, an Fe²⁺/α-ketoglutarate-dependent halogenase that dictates substrate binding order through conformational changes, a decarboxylase that establishes regiochemistry for cyclopropane formation, and a thioesterase with specificity for β-sulfated substrates that lead to terminal alkene offloading. The four curacin A pathway dehydratases reveal an intrinsic flexibility that may accommodate bulky or stiff polyketide intermediates. In the salinosporamide A pathway, active site volume determines the halide specificity of a halogenase that catalyzes for the synthesis of a halogenated building block. Structures of a number of putative polyketide cyclases may help in understanding reaction mechanisms and substrate specificities although their substrates are presently unknown.

1. Introduction

The marine environment is a rich and largely untapped source of bioactive natural products. A plethora of newly discovered compounds with an impressive array of chemical diversity presents remarkable opportunities to discover novel biosynthetic reactions and to engineer pathways.¹⁻³ Many of these interesting natural products have polyketide skeletons. Identification and sequencing of the corresponding polyketide synthase (PKS) gene clusters has led to a full understanding of the biochemistry and structural biology of many interesting enzymes. This review summarizes the current state of our knowledge of protein structures from marine PKSs.

Curacin A is a poster child for chemical diversity of marine natural products, containing a cyclopropane, thiazoline, *cis* double bond and terminal alkene (Fig. 1A). Curacin A was isolated from cultures of the marine cyanobacterium *Moorea producta* (formerly *Lyngbya majuscula*⁴) based on its cytotoxic activity.⁵ The sequence of the *cur* gene cluster followed ten years later (Fig. 1B)⁶ but was not immediately informative about the biosynthetic steps leading to all the interesting functional groups. The curacin biosynthetic gene cluster encodes eight polyketide synthase (PKS) modules, consistent with the polyketide backbone of curacin A. A nonribosomal peptide synthetase (NRPS) module incorporates and cyclizes

*Correspondence: JanetSmith@umich.edu.

cysteine, forming the curacin thiazoline moiety. Elucidation of the biosynthetic steps for the unique cyclopropane and terminal alkene emerged in a series of biochemical studies.⁷⁻¹⁰ Three-dimensional structures for nine protein domains from the Cur pathway have been published.^{7, 11-15} These proteins along with the salinosporamide chlorinase, SaL¹⁶, and three structural genomics targets (Protein Data Bank access codes: 3GZB, 3KKG and 3LZA) are the sum of our knowledge of marine polyketide synthase structural biology (Fig. 2). The structures and enzymatic reactions are not unique to the marine environment, as the proteins have homologues in natural product biosynthetic pathways from terrestrial organisms. Knowledge of these marine systems will be applied to identify, understand and engineer homologous reactions in many other natural product biosynthetic pathways.

2. Dual-function enzyme for pathway loading – CurA GNAT

Modular PKSs initiate polyketide synthesis by priming the first pathway ACP with an acyl group that lacks the carboxylate of typical polyketide building blocks. Priming typically occurs in a loading module where an acyltransferase (AT) domain selects and loads the appropriate building block, typically malonate or methylmalonate, and a “loading” ketosynthase (KS) domain decarboxylates the acyl group. The resulting acetyl- or propionyl-ACP is ready for extension in subsequent modules of the pathway. The loading KS is competent for decarboxylation but lacks a catalytic cysteine residue and thus also lacks the condensation activity of an elongating KS.

An alternative and simpler loading strategy is used by the curacin PKS (Fig. 2).⁶ Remarkably, both the loading and decarboxylation reactions occur in the active site of a “GNAT” domain in the loading module.⁷ GNATs (GCN5-related *N*-acetyltransferases) are an enzyme superfamily for acyl-group transfer from CoA to nitrogen acceptors on a variety of biomolecules.^{17, 18} The curacin GNAT is the prototype for a new branch of the GNAT superfamily that catalyzes *S*-acyltransfer from CoA to ACP. The GNAT loading strategy is not unique to the curacin PKS nor is it exclusively marine. GNAT homologues are broadly distributed in the PKS loading modules for onnamide (*Theonella swinhoei* symbiont¹⁹), apratoxin (*Lyngbya bouillonii*²⁰) and perhaps bryostatin (*Candidatus Endobugula sertula*²¹) from marine sources, and in the PKS loading modules for pederin (*Paederus fuscipes* symbiont²²), myxovirescin A (*Myxococcus xanthus*²³), batumin/kalimantacin (*Pseudomonas fluorescens*²⁴), saxitoxin (*Cylindrospermopsis raciborskii*²⁵), and rhizoxin (*Burkholderia rhizoxina*²⁶) from terrestrial sources.

The small, 200-residue CurA GNAT domain has the remarkable ability to catalyze both malonyl-CoA decarboxylation and CoA-ACP acetyltransfer within the single active site of a GNAT-like protein.⁷ CurA GNAT has an overall structure similar to other GNAT superfamily members. The active site is at the junction of two tunnels into the protein (Fig. 3A). The crystal structure of a complex with malonyl-CoA identified the entry tunnel for the CoA donor substrate. The second tunnel is matched in length and diameter to the ACP phosphopantetheine. Conserved histidine and threonine side chains at the junction of the tunnels (Fig. 3B) were shown to be essential for decarboxylation, but not for acetyl transfer. It is likely that the iso-energetic thioester exchange between CoA and ACP does not require a catalytic assist from the enzyme. However, a conserved arginine residue adjacent to the active site may facilitate acetyl transfer by stabilizing the thiolate form of the ACP phosphopantetheine.

3. Decarboxylase with exquisite regioselectivity – CurF ECH₂

Branched carbon skeletons are a principal form of polyketide chemical diversity. Branches at positions α to the thioester of the polyketide intermediate are incorporated as branched building blocks (methylmalonyl-CoA or ethylmalonyl-CoA) or through direct methylation

by C-methyltransferase domains within the PKS module. Branches at positions β to the intermediate thioester are less common. The most widely distributed system for β -branching, the “HCS cassette” (Fig. 1A,B), includes five enzymes (see Calderone 2008 for a review²⁷). The β -branching enzyme, HCS, condenses acetate derived from acetyl-ACP with the β -carbonyl of the polyketide intermediate. HCS is adapted from a primary metabolic enzyme in the isoprenoid pathway, hydroxymethylglutaryl (HMG)-CoA synthase (HCS). Conversion of the HCS product to a simple carbon branch requires the action of two enzymes of the enoyl-CoA hydratase (ECH) family whose members have similar structures and catalyze reactions with an enolate intermediate. In order, ECH₁ removes the β -hydroxyl by dehydration to create a Δ^2 (α - β) unsaturated product; then ECH₂ removes the carboxylate from the β -branch. Subsequent action of a PKS enoylreductase (ER) domain may reduce the α - β double bond. These key enzymatic activities have been confirmed for the HCS, ECH₁, ECH₂ and ER domains of several pathways.^{8, 28-30}

The crystal structure of the CurF ECH₂ (Fig. 4A)¹³ places the ECH₂ enzymes securely within the ECH family of the crotonase superfamily. The structure is consistent with the mechanism involving an enolate intermediate (Fig. 4B). The active site is located in a hydrophobic pocket with few polar groups. These include two peptide amides (Ala78 and Gly118) that form an “oxyanion hole” to stabilize the enolate intermediate. The side chain of His240 is positioned to recognize the substrate carboxyl group and assist in CO₂ release. This histidine is invariant among ECH₂ domains of HCS cassettes and is essential for decarboxylation.¹³

The β -branch can lead to considerable chemical diversity, for example cyclopropane in curacin A, vinyl chloride in jamaicamide³¹, vinyl in pederin²², and methylmethoxy in myxovirescin.²³ Some of this diversity arises from the regiochemical outcome of the decarboxylation step catalyzed by ECH₂ (Fig. 4B). While curacin ECH₂ (CurF ECH₂) produces a Δ^2 (α - β) unsaturated product, the 59% identical jamaicamide ECH₂ (JamJ ECH₂) produces a Δ^3 (β - γ) unsaturated product.⁹ This difference leads to cyclopropane formation in curacin and vinyl chloride formation in jamaicamide. The key chemical step in determining the regiochemical outcome is protonation of the enolate intermediate following decarboxylation. In the case of CurF ECH₂, protonation at C4 generates the Δ^2 (α - β) unsaturated product. The only candidate proton donors in the active site are the side chains of Tyr82 and Lys86. Lys86, but not Tyr82, is essential for decarboxylation and thus is likely the proton donor.¹³ However, Tyr82 may protect the C2 position from protonation; ECH₂ with a Y82F substitution produced in a mixture of Δ^2 (α - β) and Δ^3 (β - γ) unsaturated products.⁹ Tyr82 and Lys86 are in a hypervariable region of ECH₂ (Fig. 4A) and are not conserved. This hypervariable region, which serves as a lid over the active site, is poorly ordered in structures of many ECH family members. In ECH₂ domains within HCS cassettes the hypervariable region is tailored to the ECH₂ substrate and the regiochemistry of the product.

The quaternary structure of the ECH₂ domain is a final wrinkle on this story. The excised CurF ECH₂ domain is trimeric like virtually all members of the ECH family and the crotonase superfamily. ECH₂ is the N-terminal domain of CurF, and is followed by nine other domains. Based on structures of homologues, most of these domains are predicted to be monomeric. However, the ketosynthase (KS) domain is likely dimeric based on KS crystal structures^{32, 33}, and the dehydratase (DH) domain is a dimer in solution and in the crystal structure (Fig. 2).¹¹ Thus, the ECH₂ trimer seems incompatible with downstream dimeric domains of KS and DH.

4. Halogenating enzymes

Many polyketides from marine organisms are halogenated. Nature employs several interesting and remarkably diverse chemical strategies for halogenation.³⁴ Structures are known for two types of halogenating enzymes from marine polyketide pathways: an $\text{Fe}^{2+}/\text{O}_2/\alpha$ -ketoglutarate-dependent halogenase from the curacin pathway¹⁵ and an *S*-adenosylmethionine (SAM)-dependent halogenase from the salinosporamide A pathway.¹⁶ In addition, a flavin-dependent halogenase is proposed to incorporate bromide in an alkynyl bromide group of jamaicamide A.³¹ Alkynyl bromides were thought to be rare in nature, but were discovered recently in the veraguamide family of marine natural products.^{35, 36}

4.1. Cryptic halogenation in curacin A biosynthesis – CurA Hal

The curacin pathway includes a halogenase (CurA Hal) that catalyzes a cryptic chlorination reaction en route to the curacin cyclopropane. A related halogenase in the jamaicamide pathway (JamE Hal, 92% identical to CurA Hal) leads to vinyl chloride in its natural product. The timing of the halogenation reaction is between the HCS and ECH_1 reactions catalyzed by the HCS cassette (Fig. 1A).⁹ Hal acts specifically on the *S*-hydroxy-3-methylglutaryl-ACP ((*S*)-HMG-ACP) product of the HCS, converting it to *S*-hydroxy-3-methyl-4-chloroglutaryl-ACP (4-Cl-HMG-ACP). Following dehydration by ECH_1 and decarboxylation by ECH_2 , the chloride is removed in the cyclopropanase step catalyzed by the CurF enoylreductase (ER). In contrast, the chloride incorporated by JamE Hal is retained as vinyl chloride in jamaicamide. The difference in final products is due to the difference between the CurF and JamJ ECH_2 reactions (Δ^2 [α - β] vs. Δ^3 [β - γ] unsaturated products) outlined above.⁹

The CurA and JamE halogenases belong to the large family of Fe^{2+} , Cl^- , α -ketoglutarate (α KG) and O_2 -dependent oxygenases. They represent the third example of an Fe^{2+}/α KG-dependent halogenase in a natural product pathway; structures of the SyrB2³⁷ and CytC3³⁸ halogenase are reviewed elsewhere in this issue.³⁹ The Fe^{2+}/α KG-dependent oxygenases decarboxylate α KG using dioxygen to form a highly reactive Fe^{4+} -oxo intermediate^{40, 41}, which abstracts a hydrogen atom from the substrate ((*S*)-HMG-ACP for CurA Hal). In halogenases, the resulting substrate radical is halogenated by “rebound” of the Fe-bound chloride.^{42, 43}

Crystal structures of CurA Hal in multiple conformations revealed an α KG-mediated conformational switch.¹⁵ The structures capture the protein in distinct open and closed conformations, and in several ligand states (Fig. 5A,B). The free enzyme, lacking α KG, O_2 , Cl^- and (*S*)-HMG-ACP, is in an open conformation in which the active site lid is disordered (Fig. 5B). Intriguingly, in the open form, loops in the enzyme are collapsed over the (*S*)-HMG-ACP binding site (Fig. 5B,D). α KG binding triggers a switch to a closed (or semi-closed) form in which the (*S*)-HMG-ACP site is accessible. The trigger for the conformational switch is a salt bridge between α KG and Arg 241 (Fig. 5C,D). Such structural changes are unprecedented in the family of Fe^{2+}/α KG-dependent oxygenases. The HMG-excluded open state of CurA Hal may have arisen from a need to protect Fe^{2+} from the (*S*)-HMG-ACP carboxylate, which would be an excellent Fe^{2+} ligand, until α KG and Cl^- have occupied three positions in the Fe coordination sphere (Fig. 5C).

4.2. A halogenated building block for salinosporamide A – SalL

The proteasome inhibitor salinosporamide A from the bacterium *Salinispora tropica* has a rare and interesting chloroethylmalonyl building block.⁴⁴ SalL is a 5'-chloro-5'-deoxyadenosine (5'-CIDA) synthase, catalyzing the conversion of SAM and chloride to methionine and 5'-CIDA in the first step along a biosynthetic sequence to

chloroethylmalonyl-CoA through a chlororibose intermediate¹⁶ (Fig. 6A). This synthesis of chloroethylmalonyl-CoA begins with the deconstruction of SAM in an unusual S_N2 reaction in which a chloride nucleophile displaces methionine from SAM. Inactivation of *salL* results in the exclusive production of the non-chlorinated salinosporamide B. Supplementation of the media in a *salL*⁻ culture with 5'-CIDA restores salinosporamide A production, confirming the role of SalL as a 5'-CIDA synthase *in vivo*.¹⁶

SalL belongs to the same structural family as the bacterial fluorinase, 5'-fluoro-5'-deoxyadenosine synthase (5'-FDAS), from *Streptomyces cattleya*.⁴⁵ SalL organizes as a homotrimer with each of the three active sites positioned at a subunit interface (Fig. 2).¹⁶ In contrast to 5'-FDAS, SalL can act on a wide range of halide substrates; in addition to chloride, SalL can utilize bromide and iodide, but not fluoride, to produce 5'-BrDA and 5'-IDA, respectively. Structures were solved for SalL with the products 5'-CIDA and methionine, SalL with adenosine, SalL Y70T (an inactive variant) with substrates chloride and SAM, and SalL Y70T/G131S (a partially active variant) with 5'-CIDA and methionine.¹⁶

Differences in the active site volumes of SalL and 5'-FDAS give insight into the halide specificity of 5'-FDAS and conversely the lack of specificity of SalL. An extremely compact halide-binding pocket of 5'-FDAS excludes all halide species except desolvated fluoride. Desolvation increases halide nucleophilicity and may be required for efficient catalysis.^{16, 45} In contrast, the halide-binding pocket in SalL is too large to effectively desolvate fluoride, but accommodates larger desolvated ions, explaining the ion selectivity of this enzyme (Fig. 6B,C). The chloride ion, positioned by a hydrogen bond from the backbone amide of Gly131, is aligned almost exactly with the S^+-C5' bond, supporting a proposed S_N2 reaction mechanism (Fig 6B).

5. Presentation of substrates to the HCS cassette – tandem ACPs

The substrate (*S*)-HMG-ACP is presented to Hal by one or more of a triple ACP at the CurA C-terminus. Tandem, nearly identical, ACPs are common in HCS cassettes.²⁷ Hal activity is enhanced by each of the ACPs in the triple as well as by an 8.6-kDa C-terminal extension of CurA.⁴⁶ Electron microscopy and analytical ultracentrifugation of the triple-ACP show that the C-terminal extension mediates dimerization.⁴⁶ The ACPs are structurally independent. Solution structures of holo-ACP1 and HMG-ACP1 have the expected four-helix ACP architecture with the HMG-phosphopantetheine on a serine at the start of helix $\alpha 2$ (Fig. 7).¹² Substitutions at a few residues on ACP helices $\alpha 2$ and $\alpha 3$ near the serine attachment site reduced the effectiveness of ACP1 in delivery of substrate to CurA Hal, consistent with a specific ACP-Hal interaction.

6. Polyketide synthase dehydratases – CurF, CurH, CurJ and CurK DHs

Many marine polyketides contain exotic bulky or rigid functional groups. Such intermediates pose a problem for PKS dehydratase (DH) domains as the active site is at the center of a narrow tunnel in the protein. A convenient solution to the tunnel problem is apparent in crystal structures of four “standard” PKS DH domains from the curacin biosynthetic pathway.¹¹ Comparison of the curacin DH structures with the DH from the erythromycin pathway⁴⁷ suggests that features necessary to accommodate bulky and constrained substrates seen in marine PKS reactions are likely common to all DHs.

PKS dehydratase domains are closely related to the eukaryotic fatty acid synthase (FAS) dehydratase domains. They have a double “hot-dog” fold with a catalytic histidine contributed by the N-terminal hot dog and a catalytic aspartate contributed by the C-terminal hot dog. Both PKS and FAS DH domains are dimers and contribute to the dimer interface of

the megaenzyme complex in which they are situated. Whereas the DH modules in the FAS structures form a V-shaped dimer^{48, 49}, curacin DHs share with the erythromycin DH an extended, linear, dimeric form.^{11, 47} While it was initially surmised that the extended conformation was an artifact of the crystal form for the erythromycin DH, the observation of the same extended structure with all four curacin DH domains suggests that this is the form of the DH dimer in the megaenzyme. If so, there are likely implications for the arrangement of the subsequent ketoreductase and enoylreductase domains in PKS modules.

The active site within a DH domain is located at the crook of a deep V-shaped tunnel (Fig. 8). In bacterial FAS systems, the tunnel accommodates the fatty-acyl intermediate.⁵⁰ However, the intermediates of polyketide synthesis are frequently bulky and sterically restrained by conjugated double bonds. For example, the putative CurK DH substrate with its multiple cyclic groups and conjugated double bonds (Fig. 1A) would be unable to thread through a constricted and bent access tunnel (Fig. 8). Access to the active site is likely facilitated by movement of two structural elements that cover the tunnel: a loop and a short C-terminal α -helix. In the five PKS DH crystal structures, amino acids in the tunnel cover typically have elevated B-factors. In some structures, the covers occur in multiple conformations, demonstrating that these regions are flexible. Opening of the cover allows access to the active site and circumvents the need to thread the substrate through the tunnel.

A second consequence of the DH active site location is that the PKS intermediate must bend $\sim 90^\circ$ to access the catalytic histidine and aspartate (Fig. 8). This bend, combined with the stereochemistry at the β -OH is proposed to determine the isomeric outcome of the DH reaction (*cis* vs. *trans*).^{11, 47} The β -OH stereochemistry is set by the preceding ketoreductase reaction. The two stereochemical outcomes are classified A-type and B-type, where A-type KRs yield a β -OH stereochemistry like the (3*S*)-3-hydroxyacyl-CoA intermediates and B-type KRs give a β -OH stereochemistry like the (3*R*)-3-hydroxyacyl-CoA intermediates.^{51, 52} The sharp bend in the DH tunnel at the β -OH position is surmised to restrict the product isomer so that A-type substrates yield *cis* double-bonds and B-type substrates yield *trans* double-bonds.^{11, 47} The structures of all four curacin DH domains and their known or presumed reaction products are consistent with this hypothesis.

7. Terminal alkene formation in curacin A – CurM TE

The curacin pathway catalyzes the unprecedented offloading of an alkene. Typically, polyketide offloading by a terminal thioesterase produces a linear carboxylate or a macrolactone. In curacin A biosynthesis the combined actions of a sulfotransferase (ST) and a thioesterase (TE) in the final module (CurM) create a terminal alkene (Fig. 1). The ST sulfonates the β -hydroxyl group of the ACP-linked intermediate using the sulfonate donor 3'-phosphoadenosine 5'-phosphosulfate (PAPS).¹⁰ CurM TE then acts specifically upon the β -sulfated activated intermediate to hydrolyze the thioester and eliminate CO₂ and sulfate, yielding the alkene product.¹⁰ Both the ST and TE activities are without precedent. Biological sulfonation to activate a substrate for catalysis had not been observed previously. This is also the first observation of a thioesterase with activities of decarboxylation and sulfate elimination in addition to thioester hydrolysis.

CurM TE has many features in common with standard PKS offloading TEs, including the α / β hydrolase core and the Ser-His-Glu catalytic triad, but substantial differences adapt it to β -sulfate recognition.¹⁴ For example, the typical TE "lid" is fixed in an open position, exposing the active site (Fig. 9A). In contrast, the lid of other TEs is either permanently closed by dimer contacts, as in the case of PKS offloading TEs,^{11, 53-56} or is flexible, as in the case of editing or specialized PKS TEs and NRPS TEs.⁵⁷⁻⁶¹ The CurM TE lid is held open by extensive lid-core contacts on one side and by dimer contacts on the other (Fig.

9A,B). In contrast, dimerization of typical PKS offloading TEs fixes the lid in a closed state that forms a narrow tunnel with the active site at its center. With the open active site, one might expect CurM TE to have a very broad substrate range. However, it is in fact specific for β -sulfated substrates.

The active site environment suggests how CurM TE catalyzes the unique combined decarboxylation, β -sulfate elimination, and hydrolysis to produce the terminal alkene. The catalytic triad is located in the middle of an exposed cleft created by the open lid (Fig. 9C). Arg205, located within the cleft, is in an optimal position to recognize the substrate β -sulfate (Fig. 9D). CurM TE has ~800 fold greater hydrolysis activity for β -sulfated substrates compared to β -hydroxy substrates.¹⁰ Substitutions at Arg205 resulted in an almost complete loss of activity for a β -sulfated substrate (2-6% of wild type), but activity for a β -hydroxyl substrate was near wild-type levels.¹⁴ In the very open active site, Arg205 may provide the only interaction to position substrates for catalysis at the catalytic triad, thereby creating specificity for β -sulfate. Arg205 may also assist in stabilizing the sulfate group during the decarboxylation/sulfate elimination step.

Open reading frames in genome sequences of five cyanobacteria and two proteobacteria (gammaproteobacterium and myxobacterium) that encode ACP-ST-TE domains have substantial identity (50%-33% at the amino acid level) to CurM ACP-ST-TE (Fig. 10A). The sequences come from both marine and terrestrial sources. The TE sequences within these putative ACP-ST-TE tri-domains conserve critical features of CurM TE including the dimer interface, lid secondary structure, and active site, suggesting that they catalyze the same reactions as CurM TE with the same β -sulfate specificity. The other cyanobacterial sequences are not part of natural product biosynthetic gene clusters, but instead encode stand-alone eight-domain polypeptides that may convert long-chain fatty acids to terminal-alkene hydrocarbons due to an acyl-activating domain at the N-terminus (Fig. 10B). This possibility has been confirmed in deletion studies in one system.⁶² Thus, the curacin offloading strategy has potential application in biofuels development.

8. Orphan proteins – SnoaL-like putative cyclase

Structures of a number of putative natural product biosynthetic enzymes have been determined as part of structural genomics initiatives, and it is likely that more will follow. These include three putative SnoaL-like polyketide cyclases from *Shewanella denitrificans*, *Shewanella putrefaciens* and *Jannaschia* sp. (PDB 3GZB, 3KKG and 3LZA). SnoaL catalyzes the cyclization of nogalamycin precursors through intramolecular aldol condensation. The structures reported for SnoaL-like polyketide cyclases share essential features with SnoaL, including of active-site residues and a substrate binding pocket of similar size and shape.⁶³ However, the natural substrates for these proteins are unknown, and the ORF positions within the respective genomes are not obviously within biosynthetic gene clusters, complicating interpretations about their natural functions.

9. Conclusion

The marine environment is a rich source of polyketides and other natural products of amazing chemical diversity. We are beginning to couple the knowledge of chemical diversity with discoveries of biosynthetic pathways, biochemical functions and protein three-dimensional structures. The future is bright with a promise to discover even more interesting biosynthetic reactions and enzyme structures with great chemoenzymatic potential and as tools in pathway engineering.

References

1. Jones AC, Monroe EA, Eisman EB, Gerwick L, Sherman DH, Gerwick WH. *Nat Prod Rep*. 2010; 27:1048–1065. [PubMed: 20442916]
2. Lane AL, Moore BS. *Nat Prod Rep*. 2011; 28:411–428. [PubMed: 21170424]
3. Tan LT. *Phytochemistry*. 2007; 68:954–979. [PubMed: 17336349]
4. Engene N, Rottacker EC, Kastovsky J, Byrum T, Choi H, Ellisman MH, Komarek J, Gerwick WH. *Int J Syst Evol Microbiol*. 2011
5. Gerwick WH, Proteau PJ, Nagle DG, Hamel E, Blokhin A, Slate DL. *J Org Chem*. 1994; 59:1243–1245.
6. Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Jia J, Sherman DH, Gerwick WH. *Journal of natural products*. 2004; 67:1356–1367. [PubMed: 15332855]
7. Gu L, Geders TW, Wang B, Gerwick WH, Hakansson K, Smith JL, Sherman DH. *Science*. 2007; 318:970–974. [PubMed: 17991863]
8. Gu L, Jia J, Liu H, Hakansson K, Gerwick WH, Sherman DH. *Journal of the American Chemical Society*. 2006; 128:9014–9015. [PubMed: 16834357]
9. Gu L, Wang B, Kulkarni A, Geders TW, Grindberg RV, Gerwick L, Hakansson K, Wipf P, Smith JL, Gerwick WH, Sherman DH. *Nature*. 2009; 459:731–735. [PubMed: 19494914]
10. Gu L, Wang B, Kulkarni A, Gehret JJ, Lloyd KR, Gerwick L, Gerwick WH, Wipf P, Hakansson K, Smith JL, Sherman DH. *Journal of the American Chemical Society*. 2009; 131:16033–16035. [PubMed: 19835378]
11. Akey DL, Razelun JR, Tehranisa J, Sherman DH, Gerwick WH, Smith JL. *Structure*. 2010; 18:94–105. [PubMed: 20152156]
12. Busche A, Gottstein D, Hein C, Ripin N, Pader I, Tufar P, Eisman EB, Gu L, Walsh CT, Sherman DH, Lohr F, Guntert P, Dotsch V. *ACS Chem Biol*. 2011
13. Geders TW, Gu L, Mowers JC, Liu H, Gerwick WH, Hakansson K, Sherman DH, Smith JL. *J Biol Chem*. 2007; 282:35954–35963. [PubMed: 17928301]
14. Gehret JJ, Gu L, Gerwick WH, Wipf P, Sherman DH, Smith JL. *J Biol Chem*. 2011; 286:14445–14454. [PubMed: 21357626]
15. Khare D, Wang B, Gu L, Razelun J, Sherman DH, Gerwick WH, Hakansson K, Smith JL. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:14099–14104. [PubMed: 20660778]
16. Eustaquio AS, Pojer F, Noel JP, Moore BS. *Nat Chem Biol*. 2008; 4:69–74. [PubMed: 18059261]
17. Dyda F, Klein DC, Hickman AB. *Annu Rev Biophys Biomol Struct*. 2000; 29:81–103. [PubMed: 10940244]
18. Vetting MW, SdC LP, Yu M, Hegde SS, Magnet S, Roderick SL, Blanchard JS. *Arch Biochem Biophys*. 2005; 433:212–226. [PubMed: 15581578]
19. Piel J, Hui D, Wen G, Butzke D, Platzer M, Fusetani N, Matsunaga S. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101:16222–16227. [PubMed: 15520376]
20. Grindberg RV, Ishoe T, Brinza D, Esquenazi E, Coates RC, Liu WT, Gerwick L, Dorrestein PC, Pevzner P, Lasken R, Gerwick WH. *PLoS One*. 2011; 6:e18565. [PubMed: 21533272]
21. Sudek S, Lopanik NB, Waggoner LE, Hildebrand M, Anderson C, Liu H, Patel A, Sherman DH, Haygood MG. *Journal of natural products*. 2007; 70:67–74. [PubMed: 17253852]
22. Piel J. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99:14002–14007. [PubMed: 12381784]
23. Simunovic V, Zapp J, Rachid S, Krug D, Meiser P, Muller R. *Chembiochem*. 2006; 7:1206–1220. [PubMed: 16835859]
24. Mattheus W, Gao LJ, Herdewijn P, Landuyt B, Verhaegen J, Masschelein J, Volckaert G, Lavigne R. *Chem Biol*. 2010; 17:149–159. [PubMed: 20189105]
25. Kellmann R, Mihali TK, Jeon YJ, Pickford R, Pomati F, Neilan BA. *Appl Environ Microbiol*. 2008; 74:4044–4053. [PubMed: 18487408]
26. Partida-Martinez LP, Hertweck C. *Chembiochem*. 2007; 8:41–45. [PubMed: 17154220]

27. Calderone CT. *Nat Prod Rep*. 2008; 25:845–853. [PubMed: 18820753]
28. Buchholz TJ, Rath CM, Lopanik NB, Gardner NP, Hakansson K, Sherman DH. *Chem Biol*. 2010; 17:1092–1100. [PubMed: 21035732]
29. Calderone CT, Iwig DF, Dorrestein PC, Kelleher NL, Walsh CT. *Chem Biol*. 2007; 14:835–846. [PubMed: 17656320]
30. Calderone CT, Kowtoniuk WE, Kelleher NL, Walsh CT, Dorrestein PC. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:8977–8982. [PubMed: 16757561]
31. Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA, Gerwick WH. *Chem Biol*. 2004; 11:817–833. [PubMed: 15217615]
32. Tang Y, Kim CY, Mathews II, Cane DE, Khosla C. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:11124–11129. [PubMed: 16844787]
33. Tang Y, Chen AY, Kim CY, Cane DE, Khosla C. *Chem Biol*. 2007; 14:931–943. [PubMed: 17719492]
34. Neumann CS, Fujimori DG, Walsh CT. *Chem Biol*. 2008; 15:99–109. [PubMed: 18291314]
35. Mevers E, Liu WT, Engene N, Mohimani H, Byrum T, Pevzner PA, Dorrestein PC, Spadafora C, Gerwick WH. *Journal of natural products*. 2011; 74:928–936. [PubMed: 21488639]
36. Salvador LA, Biggs JS, Paul VJ, Luesch H. *Journal of natural products*. 2011; 74:917–927. [PubMed: 21446699]
37. Blasiak LC, Vaillancourt FH, Walsh CT, Drennan CL. *Nature*. 2006; 440:368–371. [PubMed: 16541079]
38. Wong C, Fujimori DG, Walsh CT, Drennan CL. *Journal of the American Chemical Society*. 2009; 131:4872–4879. [PubMed: 19281171]
39. Drennan CL. *Journal of natural products*. 2012; XX:XXX–XXX.
40. Galonic DP, Barr EW, Walsh CT, Bollinger JM Jr, Krebs C. *Nat Chem Biol*. 2007; 3:113–116. [PubMed: 17220900]
41. Hoffart LM, Barr EW, Guyer RB, Bollinger JM Jr, Krebs C. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:14738–14743. [PubMed: 17003127]
42. Matthews ML, Krest CM, Barr EW, Vaillancourt FH, Walsh CT, Green MT, Krebs C, Bollinger JM. *Biochemistry*. 2009; 48:4331–4343. [PubMed: 19245217]
43. Matthews ML, Neumann CS, Miles LA, Grove TL, Booker SJ, Krebs C, Walsh CT, Bollinger JM Jr. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:17723–17728. [PubMed: 19815524]
44. Gulder TA, Moore BS. *Angew Chem Int Ed Engl*. 2010; 49:9346–9367. [PubMed: 20927786]
45. Dong C, Huang F, Deng H, Schaffrath C, Spencer JB, O'Hagan D, Naismith JH. *Nature*. 2004; 427:561–565. [PubMed: 14765200]
46. Gu L, Eisman EB, Dutta S, Franzmann TM, Walter S, Gerwick WH, Skiniotis G, Sherman DH. *Angew Chem Int Ed Engl*. 2011; 50:2795–2798. [PubMed: 21387490]
47. Keatinge-Clay A. *J Mol Biol*. 2008; 384:941–953. [PubMed: 18952099]

Fig. 1.

Curacin A: prototype for the rich biosynthesis of marine polyketides.

(A) Curacin A biosynthesis with emphasis on reactions informed by protein structures.

Changes to the polyketide intermediate are highlighted in red for each step. The molecular structures of curacin A and jamaicamide A from two strains of *Moorea producta* are shown.

Abbreviations are: ACP, acyl carrier protein; GNAT, GCN5-related *N*-acetyltransferase; KS, ketosynthase; AT, acyltransferase; HCS, HMG-CoA synthase; Hal, halogenase; ECH1, dehydrating enoyl-CoA hydratase; ECH2, decarboxylating enoyl-CoA hydratase; ER, enoyl reductase; DH, dehydratase; ST, sulfotransferase; TE, thioesterase; PAPS, 3'-phosphoadenosine-5'-phosphosulfate; PAP, 3'-phosphoadenosine-5'-phosphate. (B) Curacin A megasynthase. The annotated domains within the thirteen polypeptides (CurA – CurM) with domains are colored by type. Asterisks indicate domains with published structures. Domains not listed in (A) are: AR, GNAT adaptor region; KRC, ketoreductase catalytic domain, KRS, ketoreductase structural domain; CMT, *C*-methyl transferase; OMT, *O*-methyl transferase; A, adenylation; Cy, condensation/cyclization; PCP, peptidyl carrier protein. Box sizes are proportional to domain sizes.

Fig. 2.

Gallery of protein structures from marine PKS pathways. Each protein is shown as a ribbon diagram in its natural multimer state with subunits in contrasting colors. Substrates, cofactors or relevant side chains are rendered as spheres to emphasize the active site location in each protein. Protein Data Bank (PDB) accession codes are: CurA GNAT – 2REE, 2REF; CurA Hal – 3NNF, 3NNJ, 3NNL, 3NNM; CurA ACP – 2LIW; CurF ECH₂ – 2Q2X, 2Q34, 2Q35; CurF DH – 3KG6; CurH DH – 3KG7; CurJ DH – 3KG8; CurK DH – 3KG9; CurM TE – 3QIT; SalL – 2Q6K, 2Q6I, 2Q6L, 2Q6O; SnoaL-like – 3GZB, 3KKG, 3LZA.

Fig. 3. CurA GNAT loading enzyme.⁷ **(A)** Substrate and product tunnels through GNAT. The lower tunnel (tan surface) is the entrance for the malonyl-CoA substrate, based on electron density from crystals soaked in malonyl-CoA. The upper tunnel (gold surface) is for the ACP phosphopantetheine acceptor. Acetyl-CoA and the phosphopantetheine of ACP are shown in stick form with black C atoms. Side chains of key amino acids (Trp249, Thr355, His389 and Arg404) are shown in stick form with yellow C atoms. **(B)** Close-up view of the active site. Thr355 and His389 are essential for decarboxylation but not acetyl transfer activity. Hydrogen bonds from these residues to the acetyl carbonyl are shown (dashed lines). Trp249 constricts the tunnel at the active site. Arg404 may stabilize the reactive thiolate form of ACP-ppant.

Fig. 4. CurF ECH₂ decarboxylase.¹³ **(A)** ECH₂ with modeled substrate, 4-chloro-3-methylglutaconyl-ACP. The substrate (gray C atoms), and key amino acid side chains (yellow C atoms) are shown as sticks and the protein as a green ribbon with the hypervariable region in magenta. The substrate thioester O atom is hydrogen bonded (dashed lines) to protein amides of the ECH₂ oxyanion hole. Invariant His240 is essential for catalysis, as is Lys86 on the hypervariable helix α2. **(B)** Proposed reaction mechanism for ECH₂ decarboxylation. Following decarboxylation, Lys86 is proposed to donate a proton to the C4 atom to generate the Δ² (α-β) unsaturated product of CurF ECH₂, 4-chloro-3-methylcrotonyl-ACP, which is subsequently converted to the cyclopropyl moiety of curacin A. In contrast, the ECH₂ of the jamaicamide pathway donates a proton to the C2 atom to generate the Δ³ (β-γ) unsaturated product and the vinyl chloride is retained in jamaicamide.

Fig. 5. Conformational states of CurA Hal.¹⁵ **(A)** Closed form Hal. A lid (red) covers the active site and a loop (grey) connecting $\alpha 9$ - $\alpha 10$ is away from the active site. **(B)** Open form Hal. The lid between $\alpha 2$ and $\alpha 4$ is disordered and the loop (grey) connecting $\alpha 9$ - $\alpha 10$ is next to the active site. α -KG is absent in the open state. **(C)** CurA Hal active site with a fully occupied iron coordination sphere. The iron center has octahedral coordination with only two protein ligands (His115 and His228). The cosubstrate, α -KG, chelates iron in a bidentate fashion. Formate from the crystallization solution occupies the sixth coordination position, which is thought to contain water in the resting enzyme and dioxygen during catalysis. **(D)** Conformational switch in CurA Hal. Upon departure of α -KG, the Arg241 side chain swings out, initiating a cascade of events. Strand $\beta 3$ shifts towards strand $\beta 2$, which becomes disordered. Tyr68 moves out of the active site. Asp283 moves towards the active site where it forms a salt bridge with the Arg247 side chain, which has moved out of the active site. His99 swings away from the active site to hydrogen bond with Asp69. Closed and open conformations are depicted in cyan and yellow, respectively.

Fig. 6. SalL chlorinase.¹⁶ (A) Chloroethylmalonate building block. SalL catalyzes the conversion of chloride and SAM to 5'-CIDA and methionine. 5'-CIDA is further processed to yield chloroethylmalonyl-CoA, which is incorporated as a building block in the salinosporamide PKS pathway. (B) Composite image of the active site from wild type SalL (PDB 2Q6I) with chloride and SAM substrates from SalL/Y70T/G131S (PDB 2Q6O) shown in cartoon, and (C) in the same orientation, surface view.

Fig. 7. Solution structure of HMG-ACP1.¹² The HMG-Ppant arm is linked to the Ser1989 side chain. Substitutions in $\alpha2$ (Asp1988, Ile1990) and $\alpha3$ (Ala2009) reduce chlorination efficiency, implicating a role for these residues in CurA Hal recognition.

Fig. 8. Curacin dehydratase.¹¹ (A) Top view of CurK DH active site surface after removal of the proposed mobile elements (blue ribbon). The CurK DH substrate (yellow) is modeled with double bonds explicitly indicated. The active site His/Asp dyad is indicated with blue and red patches on the protein surface. (B) Cut-away view of (A) from the side. The V-shaped trench imposes restrictions on substrate placement. Active site His and Asp are shown in stick representation.

Fig. 9. CurM TE decarboxylation thioesterase.¹⁴ **(A)** CurM TE monomer. The catalytic triad active site (magenta sticks) is located in a cleft between the core (green ribbon) and open lid (purple ribbon). The lid is held in place by specific contacts of Arg185 in the lid and Asp57 in the core. **(B)** CurM TE dimer. The novel lid (purple) to core (green) dimer interface holds the lid in an open position. Extensive subunit contacts are apparent in the translucent surface rendering, colored as in (A). **(C)** Active site cleft in CurM TE monomer. The acyl enzyme intermediate (yellow C atoms) is modeled into the open cleft between the lid (purple) and core (green). **(D)** CurM TE active site showing the catalytic triad (Ser100, His266, Glu124), sulfate-recognizing Arg205, and the modeled acyl enzyme intermediate (yellow C atoms).

Fig. 10.

Distribution of CurM TE homologues. **(A)** ORFs encoding ACP-ST-TE domains. The predicted domains are depicted for all of the sequences encoding sequential ACP-ST-TE domains. GenBank entries are: *Moorea Producta* (ACV42478), *Synechococcus* PCC 7002 (YP_001734428), *Cyanothece* PCC 7424 (YP_002377174), *Cyanothece* PCC 7822 (ZP_03153601), *Moorea Producta* 3L (ZP_08425908), *Prochloron didemni* (AEH57210), *Pseudomonas entomophila* L48 (YP_610919), *Haliangium ochraceum* DSM 14365 (YP_003265308). **(B)** Predicted pathway to hydrocarbons based on sequence annotation. The acyl-activating (AA) domain is predicted to load a fatty acid onto the adjacent ACP, the KS-AT-KR to extend and reduce the acyl-ACP to produce a β -hydroxyl substrate for the ST-TE.