Main Manuscript for

Bioinformatic and mechanistic analysis of the palmerolide PKS-NRPS biosynthetic pathway from the microbiome of an Antarctic ascidian

Nicole E. Avalon;[1] Alison E. Murray;[2] Hajnalka E. Daligault;[3] Chien-Chi Lo;[3] Karen W. Davenport;[3] Armand E.K. Dichosa;[3] Patrick S.G. Chain;[3] Bill J. Baker[1]

[1]Department of Chemistry, University of South Florida, Tampa, FL 33620

[2]Division of Earth and Ecosystem Sciences, Desert Research Institute, Reno, NV 89512

[3]Los Alamos National Laboratory, Los Alamos, NM 87545

Corresponding author:

Bill J. Baker, Department of Chemistry, University of South Florida, Tampa, FL 33620, USA. +1-813-974-1967, bjbaker@usf.edu

Author Contributions: This work was the result of a team effort in which the following contributions are recognized: conceptualization, A.E.M., P.S.G.C., and B.J.B.; methodology and experimentation, N.E.A., A.E.M., H.E.D., C.L., P.S.G.C., B.J.B; data analysis, N.E.A., A.E.M., H.E.D, C.L., K.W.D., A.E.K.D., P.S.G.C, B.J.B; data curation, N.E.A., A.E.M., H.E.D, C.L., P.S.G.C, B.J.B; original draft preparation, N.E.A., B.J.B., A.E.M., H.E.D.; review and editing, N.E.A., B.J.B., A.E.M., C.L., H.E.D., P.S.G.C., A.E.K.D; funding acquisition, A.E.M., P.S.G.C., and B.J.B. All authors have read and agreed to the published version of the manuscript.

Competing Interest Statement: The authors have no competing interests to disclose.

Classification: Biological sciences, environmental sciences and ecology

Keywords: microbial biosynthesis, secondary metabolites, Antarctic microbiology, *trans*-AT Type I polyketide synthase, marine natural products

This PDF file includes:

Main Text.
Figures 1 to 4.

Abstract

Polyketides are a complex family of natural products that often serve competitive or pro-survival purposes but can also demonstrate bioactivity in human diseases as, for example cholesterol lowering agents, anti-infectives, or anti-tumor agents. Marine invertebrates and microbes are a rich source of polyketides. Palmerolide A, a polyketide isolated from the Antarctic ascidian *Synoicum adareanum,* is a vacuolar-ATPase inhibitor with potent bioactivity against melanoma cell lines. The biosynthetic gene clusters (BGC) responsible for production of secondary metabolites are encoded in the genomes of the producers as discrete genomic elements. A putative palmerolide BGC was identified from a *S. adareanum* metagenome based on a high degree of congruence with a chemical structure-based retrobiosynthetic prediction. Protein family homology analysis, conserved domain searches, and active site and motif identification were used to identify and propose the function of the 75 kb *trans*-acyltransferase (AT) polyketide synthase-non-ribosomal synthase (PKS-NRPS) domains responsible for the synthesis of palmerolide A. Though PKS systems often act in a predictable co-linear sequence, this BGC includes multiple *trans*-acting enzymatic domains, a non-canonical condensation termination domain, a bacterial luciferase-like monooxygenase (LLM), and multiple copies found within the metagenome-assembled genome (MAG) of *Candidatus* Synoicohabitans palmerolidicus. Detailed inspection of the five highly similar pal BGC copies suggests the potential for biosynthesis of other members of the palmerolide chemical family. This is the first delineation of a biosynthetic gene cluster from an Antarctic species. These findings have relevance for fundamental knowledge of PKS combinatorial biosynthesis and could enhance drug development efforts of palmerolide A through heterologous gene expression.

Significance Statement

Complex interactions exist between microbiomes and their hosts. Increasingly, defensive metabolites that have been attributed to host biosynthetic capability are now being recognized as products of associated microbes. These unique metabolites often have bioactivity in targets of human disease and can be purposed as pharmaceuticals. The molecular machinery for production of palmerolide A, a macrolide that is potent and selective against melanoma, was discovered as a genomic cluster in the microbiome of an Antarctic ascidian. Multiple non-identical copies of this genomic information provide clues to differences in specific enzymatic domains and point to Nature's ability to perform combinatorial biosynthesis *in situ*. Harnessing this genetic information may pave a path for development of a palmerolide-based drug.

Introduction

Marine invertebrates such as corals, sponges, mollusks, and ascidians, are known to be a rich source of bioactive compounds (1). Due to their sessile or sluggish nature, chemical defenses such as secondary metabolites are often key to their survival. Many compound classes are represented among benthic invertebrates including terpenes, nonribosomal peptide synthase (NRPS) products, ribosomally synthesized and post-translationally modified peptides (RiPPs), and polyketides. It is estimated that that over 11,000 secondary metabolites from marine and

terrestrial environments understood to be of polyketide synthase (PKS) and NRPS origin have been isolated and described (2). Biosynthetic gene clusters (BGCs) exist as a series of genomic elements that encode for megaenzymes responsible for production of these secondary metabolites. BGCs can have distinct nucleotide composition properties such as codon usage and guanine-cytosine content that do not match the remainder of the genome (3, 4), suggesting a mechanism of transfer from organisms that are only distantly related, perhaps even from interkingdom horizontal gene transfer (5, 6). Interestingly, the BGCs for many natural products isolated from marine invertebrates are found in the host-associated microbiota, reflecting the role of these compounds in symbiosis (7).

Polyketides are a complex family of natural products produced by a variety of PKS enzymes that are related to, but evolutionarily divergent from, fatty acid synthases (8). They often possess long carbon chains with varied degrees of oxidation, can contain aromatic components, and may be either cyclic or linear. It is estimated that of the polyketides that have been isolated and characterized, 1% have potential biological activity against human diseases, making this class of compounds particularly appealing from a drug discovery and development standpoint (9). This potential for use as pharmaceuticals is approximately five times greater than for compounds of all other natural product classes (9). Many polyketides are classified as macrolides, which are large-ring lactones that are pharmaceutically relevant due to a number of biological actions, including, for example, targeting the cytoskeleton, ribosomal protein biosynthesis, and vacuolar type V-ATPases (10–13). V-ATPases are responsible for acidification of cells and organelles via proton transport across membranes, including those of lysosomes, vacuoles, and endosomes. These enzymes appear to have an impact on angiogenesis, apoptosis, cell proliferation, and tumor metatastisis (10). A number of marine macrolides inhibit V-ATPases, including lobatamides, chondropsins, iejimalides, and several of the palmerolides (14–18).

There are three types of PKS systems. Type I PKS systems in bacteria are primarily comprised of non-iteratively acting multimodular enzymes that lead to progressive elongation of a polyketide chain, though these megaenzymes can also include “stuttering” modules that may act iteratively (19–21). In addition, some bacterial Type I PKS systems are comprised solely of iteratively acting monomodular enzymes that catalyze a series of chain elongation steps for polyketide formation (22). Type II PKS systems typically contain separate, iteratively acting enzymes that biosynthesize polycyclic aromatic polyketides, while Type III PKS systems possess iteratively-acting homodimeric enzymes that often result in monocyclic or bicyclic aromatic polyketides (19). Type I PKS systems can be subdivided into two groups, depending upon whether the acyl transferase (AT) modules are encoded within each module at the site that is parallel to the functional role of the ATs, as in *cis*-AT Type I PKS, or physically distinct from the megaenzyme, as in *trans*-AT Type I PKS. In both cases, there are often parallel relationships between the genome order, the action of enzymatic modules, and the functional groups present in the growing polyketide chain, though in *trans*-AT systems deviations from these parallel relationships is more likely to be observed (23). In *trans*-AT systems, AT domains may be incorporated in a mosaic fashion through horizontal gene transfer (23). This introduces greater molecular architectural diversity over evolutionary time, as one clade of *trans*-ATs may select for a malonyl-CoA derivative, while the *trans*-AT domains in another clade may select for unusual or functionalized subunits (24, 25). Additionally, recombination, gene duplication, and conversion events can lead to further diversification of the resultant biosynthetic machinery (26). Predictions regarding the intrinsic relationship between a secondary metabolite of interest, the biosynthetic megaenzyme, and the biosynthetic gene cluster (BGC) can be harnessed for natural product discovery and development (27–29).

In the search for new and bioactive chemotypes as inspiration for the next generation of drugs, underexplored ecosystems hold promise as biological and chemical hotspots (30). The vast Southern Ocean comprises one-tenth of the total area of Earth’s oceans and is largely unstudied for its chemodiversity. The coastal marine environment of Antarctica experiences seasonal extremes in, for example, ice cover, light field, and food resources. Taken with the barrier to migration imposed by the Antarctic Circumpolar Current and the effects of repeated

glaciation events on speciation, a rich and endemic biodiversity has evolved, with consequent potential for new chemodiversity (30–32).

Palmerolide A (**Fig. 1**) is the principal secondary metabolite isolated from *Synoicum adareanum*, an ascidian which can be found in abundance at depths of 10 to 40 m in the coastal waters near Palmer Station, Antarctica (17). Palmerolide A is a macrolide polyketide that possesses potent bioactivity against malignant melanoma cell lines while demonstrating minimal cytotoxicity against other cell lines (17). The NCI's COMPARE algorithm was used to correlate experimental findings with a database for prediction of the biochemical mechanism by identifying the mechanism of action of palmerolide A as a vacuolar-ATPase (V-ATPase) inhibitor (33). Downstream effects of V-ATPase inhibition include an increase in both hypoxia induction factor-1α and autophagy (17, 34). Increased expression of V-ATPase on the surface of metastatic melanoma cells (34) perhaps explain palmerolide A's selectivity for UACC-62 cell lines over all cell types (17). Despite the relatively high concentrations of palmerolide A in the host tissue (0.49–4.06 mg palmerolide A x g^{-1} host dry weight) (35), isolation of palmerolide A from its Antarctic source in mass sufficient for drug development it is not ecologically nor logistically feasible. Synthetic strategies for the synthesis of palmerolide A have been reported (6–13), though a clear pathway to quantities needed for drug development has been elusive. Therefore, there is substantial interest in identifying the biosynthetic gene cluster (BGC) responsible for palmerolide A production.

Our approach to identify the palmerolide BGC (*pal* BGC) began with the characterization of the host-associated microbiome (43). The core microbiome for *Synoicum adareanum*, a persistent cohort of bacteria present in many individual ascidians, was established through analysis of occurrence of distinct amplicon sequence variants (ASV) from the iTag sequencing of the Variable 3-4 regions of the bacterial 16S rRNA (35). This ultimately led to the evaluation of the metagenome and then the assembly of the metagenome assembled genome (MAG) of *Candidatus* Synoicohabitans palmerolidicus, a verrucomicrobium in the family *Opitutaceae* (44). Contained within the genome are five non-identical copies of the *pal* BGC, of which, three correlate with palmerolides, the structures of which have been previously discovered in this macrolide family. Here we report a detailed analysis of the *pal* BGCs.

Results and Discussion

A. Retrobiosynthetic Scheme for Palmerolide A

A retrobiosynthetic scheme of the *pal* BGC was developed based on the chemical structure for palmerolide A, including modules consistent with a hybrid PKS-NRPS with tailoring enzymes for key functional groups (**Fig. 1**). We hypothesized that the initial module would be PKS-like in nature to utilize 3-methylcrotonic acid as the starter unit followed by a NRPS domain for the incorporation of glycine. PKS elongation was predicted to be an 11-step sequence resulting in 22 contiguous carbons. Modifying enzymes that are encoded co-linearly were predicted to create the architectural diversity with olefin placement, reduction of certain carbonyl groups to secondary alcohols, and full reduction of other subunits. In addition, incorporation of methylmalonyl CoA or enzymatic activity of carbon methyltransferases (cMTs) were predicted to be responsible for the placement of methyl groups C-26 and C-27 from *S*-adenosylmethionine (SAM).

Several key structural features proposed to result from the action of *trans*-acting enzymes are present. As seen in the kalimantacins (45), the carbamate on C-11 was hypothesized to be installed by a carbamoyl transferase (CT). The C-25 methyl group is located on C-17, in the β-position. This structural feature suggests the origin of this branch from hydroxymethylglutaryl-CoA Synthase (HCS) catalysis, rather than the SAM-mediated methylation occurring at the α-position, which appears to be the origin of the C-26 and C-27 methyl groups. Methylation in this acetate carboxyl position is unusual, but represented in a number of notable BGC's, such as those of the jamaicamides, bryostatins, curacin A, oocydin, pederin, and psymberin, among others; in biochemically characterized Type I PKS BGCs, HCS-mediated β-branch formation is the common mechanism (46–50). The hydroxy group on C-10, in the α-position of the acetate subunit, was

hypothesized to result from elongation resulting from hydroxymalonyl-CoA incorporation or from the action of a hydroxylase at a later stage of biosynthesis.

B. Proposed Biosynthesis and Architecture of the Putative *pal* Biosynthetic Gene Cluster

Synoicum adareanum collected near Palmer Station, Antarctica was processed to separate eukaryotic from prokaryotic cells of the host microbiome. Iterations of metagenomic analysis (44) yielded a large number of BGCs when analyzed via the antiSMASH annotation platform (51). A candidate cluster with multiple copy numbers was identified and has excellent congruence with retrobiosynthetic predictions, including the key genomic markers outlined above that are likely involved in biosynthesis of palmerolide A. Integration of BGC annotations with information from protein family homology analysis, conserved domain searches, active site and motif identification, supported the hypothesis that this putative 75 kb BGC is responsible for palmerolide A production. The core scaffold was explained by the NRPS and *trans*-AT PKS hybrid system. In addition, each of the tailoring enzymes that are expected for the distinct chemical features are encoded in the genome. The proposed BGC for palmerolide A is comprised of 14 core biosynthetic modules and 25 genes in a single operon of 74,655 bases (**Fig. 2**). Fourteen modules are co-linear and two *trans*-AT domains (modules 15 & 16) follow the core biosynthetic genes. Additional *trans*-acting genes contribute to backbone modifications with at least one gene contributing to post-translational tailoring (**Fig. 3**).

An unusual starter unit and NRPS domains of *palA.* Bioinformatic analysis of the gene sequence suggests that the initial core biosynthetic domains of *palA* (modules 1 and 2) encode for the requisite acyl carrier proteins (ACP) (**Fig. 2**). Encoded in module 1, there are three ACPs in tandem, which could serve to promote an increase in metabolite production (52). The second in series is an ACP-β containing the conserved domain sequence GXDS (53) and is likely the acceptor of a starter unit containing a β-branch, consistent with the proposed starter unit for palmerolide A, 3-methylcrotonic acid. While both *trans*-acting ATs, PalE and PalF, (**Fig. 2**) possess the catalytic active site serine which is key for the proper positioning of subunit within the hydrophobic cleft of the active site (8, 54), only the first AT, PalE, has a characteristic motif that includes an active site phenylalanine, conferring specificity for malonate selection (55). The AT selecting the methylcrotonic acid starter unit is likely the second of the two *trans*-AT domains (PalF), which lacks definitive specificity for malonyl-CoA. In support, some *trans*-acting ATs have demonstrated affinity for a wider range of substrates than their *cis*-acting counterparts (26, 56). 3-Methylcrotonyl-CoA is an intermediate of branched-chain amino acid catabolism in leucine degradation; intermediates of which can be diverted to secondary metabolite production (57). Interestingly, an upregulation of leucine catabolic pathways can be seen as a mechanism of cold adaption to maintain cell wall fluidity via production of branched chain fatty acids, a factor which may be at play in this Antarctic microorganism (57, 58). The subsequent NRPS module containing condensation (C) and adenylation (A) domains as well as a carrier protein comprise module 2. Signature sequence information and NRPSPredictor2 method (51, 59) of the A domain are consistent with selection of a glycine residue. These domains incorporate the amino acid residue, resulting in the addition of a nitrogen and two carbons in this step of the biosynthesis of palmerolide A.

In a non-canonical fashion, the carrier proteins flanking the NRPS modules do not appear to be the expected ACP and PCP for module 1 and 2, respectively. The carrier protein following the KS domain in module 1 was initially annotated as a non-β-branching ACP; however, phylogenetic analysis with the amino acid sequences of carrier proteins from other hybrid PKS-NRPS systems, demonstrates that this carrier protein is in the same clade as peptidyl-carrier proteins (***SI Appendix*, Fig. S1**). The carrier protein associated with module 2 which was initially annotated simply as a phosphopantetheine attachment site (Pfam00550.24) is found to be more phylogenetically-related to ACPs within PKS-NRPS systems (***SI Appendix*, Fig. S1**). Notably, it possesses the (D/E)xGxDSL motif for phosphopantetheine arm attachment (60) with the exception of an isoleucine rather than leucine in the final position of the motif, a residue seen in

239 this position in other ACPs from hybrid PKS-NRPS systems (***SI Appendix*, Fig. S2**). Typically, a
240 PCP would follow the domains in NRPS-like modules; however, there are exceptions in the
241 literature. For example, the BGCs for both corallopyronin and oxazolamycin contain ACPs
242 following an A domain (61, 62). This non-canonical finding could point to the acquisition of these
243 domains over time, as the carrier protein for module 1 is encoded in *palA*, the same gene as the
244 proteins for both modules 1 and 2, whereas the carrier protein for module 2 is encoded at the
245 beginning of *palB*, a gene which encodes for only PKS domains (**Fig. 2**).
246
247 **Contiguous PKS chain and *trans*-acting enzymes at site of action for *palB – palD*.** The
248 contiguous carbon backbone of palmerolide A arises from 11 cycles of elongation from a series of
249 modules with a variety of enzymatic domains that include an ACP, KS, and associated genes that
250 establish the oxidation state of each subunit (**Fig 2**). The first module of *palB* (module 3) includes
251 a dehydratase (DH) and cMT domains, a sequence which results in a chain extension
252 modification to an α,β-unsaturated thioester, a result of the action of the encoded DH. The
253 expected KR domain that would be responsible for the Δ^{22} olefin (**Fig. 2**) is not present. The
254 BGCs for bryostatin 1, corallopyronin, and sorangicin also lack an accompanying KR domain to
255 work in concert with an encoded DH. The unaccompanied DH in the first two systems are
256 deemed inactive; however, an olefin results from the DH in the absence of an accompanying KR
257 in both modules 9 and 11 of the sorangicin BGC (46, 61, 63). The subsequent cMT methylation is
258 consistent with an *S*-adenosylmethionine (SAM)-derived methyl group, as expected for C-27 in
259 palmerolide A. Module 4, spanning the end of *palB* and beginning of *palC*, includes a DH, a
260 ketoreductase (KR), and another cMT domain. This cluster of domains is predicted to result in the
261 methyl-substituted conjugated diene of the macrolide tail (C-19 through C-24, C-26 on
262 palmerolide A).
263 The substrate critical for macrolactonization of the polyketide is the C-19 hydroxy group, a
264 result of the KS and KR domains encoded in module 5 (**Fig. 2**). Additionally, a domain initially
265 annotated as a dehydratase (DHt) at this location may contribute to the final cyclization and
266 release of the molecule from the megaenzyme by assisting the terminal C domain with ring
267 closure (53). This sequence shows some homology with condensation domains and does not
268 possess the hotdog fold that is indicative of canonical dehydratases, and therefore, may not truly
269 represent a DH (64). Alternatively, this domain could be responsible for the olefin shifts to the β,γ-
270 positions, as seen in bacillaene and ambruticin biosynthesis (65, 66).
271 Only an enoyl-CoA hydratase (ECH) in addition to standard ACP and KS is encoded in
272 module 6 which would lead to a ketone function; based on our retrobiosynthetic analysis, the
273 resulting ketone at C-17 is the necessary substrate for HCS-catalyzed β-branch formation,
274 resulting in the C-25 methyl group on C-17. The HCS cassette (PalK through PalO) is comprised
275 of a series of *trans*-acting domains, including an ACP, an HCS, a free KS, and 2 additional ECH
276 modules (**Fig. 2**). The HCS cassette can act while the elongating chain is tethered to an ACP
277 module, rather than after cyclization and release (67, 68). The two ECHs in the HCS cassette
278 along with the ECH encoded in-line with the core biosynthetic genes would be responsible for
279 isomerization of a terminal methylene to the observed internal olefin. An HCS cassette formed by
280 the combination of a *trans*-KS and at least one ECH module with an HCS domain is reported in
281 several other bacterial BGCs such as bryostatin 1, calyculin A, jamaicamide, mandelalide,
282 phormidolide, and psymberin (47, 50, 53, 69–71). The domain structure for the HCS cassettes
283 has a remarkably high degree of synteny across these diverse BGCs (72).
284 There is substantial similarity in the domain structure of module 7, module 10, and module
285 13, which each include a KS, DH, and KR (**Fig. 2**). The olefin that arises from the action of
286 module 7, concomitant with carbon chain elongation, is conjugated with the Δ^{16} olefin adjacent
287 the C-17 β-branch. Modules 10 and 13 have similar enzymatic composition to 7 and are likely
288 responsible for Δ^{8} and Δ^{2} olefins. The combination of KR and DH domains are also found in
289 modules 8 and 12; however, in concert with an unidentified *trans*-acting enoyl reductase domain
290 (ER), these olefins would be reduced to fully saturated monomeric subunits. There are some
291 examples of *trans*-acting ER domains carrying out this function, including OocU in oocydin, SorN
292 in sorangicin, and MndM in mandelalide (49, 63, 71), while in other systems, such as

293 corallopyronin and leinamycin, the reductions of the olefins are largely unexplained (61, 73). The
294 reduction by a *trans*-acting enzyme often occurs while the elongating polyketide is tethered to the
295 megaenzyme, as evidenced by the downstream specificity of the KS module for Claisen-type
296 condensation with subunits containing single or double bonds (63).
297 The genetic architecture for the biosynthesis of two functional groups, essential for bioactivity,
298 is encoded in module 9 (**Fig. 2**); structure-activity relationship studies demonstrate the
299 importance of the C-10 hydroxyl group and the C-11 carbamate (74). The KR domain that is
300 present, predicting the C-11 alcohol function, serves as the substrate for the carbamoyl
301 transferase (*palQ*) in a post-translational modification (75–77). Intriguingly, a domain annotated
302 as a luciferase-like monooxygenase (LLM) initially seems out of place; however, palmerolide A
303 has a hydroxy group at C-10, which is the α-position of the acetate subunit inserted in module 9.
304 LLMs associated with BGCs may not serve as true luciferases, but instead demonstrate oxidizing
305 effects on polyketides and peptides without evidence of corresponding bioluminescence (78, 79).
306 For example, there is an overrepresentation of LLMs in *Candidatus* Entotheonella BGCs without
307 known bioluminescence (80). As demonstrated through individual inactivation of the LLM in the
308 BGC of mensacarin, a Type II PKS system, Msn02, Msn04, and Msn08 have key activity as
309 epoxidases and hydroxlases (79). There are several examples of LLMs in modular Type I PKS
310 systems. OnnC from onnamide and NazB from nazumamide are two LLMs in *Candidatus*
311 Endotheonella that are proposed to serve biosynthetically as hydroxylases (80). In calyculin and
312 mandelalide, the CalD and MndB LLMs catalyze chain shortening reactions through α-
313 hydroxylation and Baeyer-Villiger-type oxidation reactions (69, 71). Phormidolide has a LLM that
314 adds a hydroxy group which is then hypothesized to attack an olefin through a Michael-type
315 addition for cyclization with enzymatic assistance from a pyran synthase (53). The hydroxylation
316 that is key in cyclization of oocydin A is likely installed by OocK or OocM, flavin-dependent
317 monooxygenases that are contiguous to the PKS genes and are thought to act while the
318 substrate is bound to a portion of the PKS megaenzyme (49). It is this hydroxylase activity that
319 we propose for the LLM in module 9. Since the producing bacteria is yet to be cultured, it is not
320 established whether this LLM may also serve a role in bioluminescence and/or quorum sensing.
321 Further evidence for the role of the LLM is provided through alignment against other LLM's. The
322 *pal* BGC contains both LuxR family transcriptional regulator as well as the DNA-binding response
323 regulator. In addition to the annotation within Pfam00296, which includes the bacterial LLMs, the
324 sequence is homologous with the multiple sequence alignments and the hidden Markov models
325 of the TIGR subfamily 04020 of the conserved protein domain family cl19096, which is noted for
326 natural product biosynthesis LLMs (80). The subfamily occurs in both NRPS and PKS systems as
327 well as small proteins with binding of either flavin mononucleotide or coenzyme F420. Alignment
328 of the LLMs from multiple PKS systems, including palmerolide A, shows homology with model
329 sequences from the TIGR subfamily 04020 (***SI Appendix*, Fig. S3**).
330 The addition of C-5 and C-6 and the reduction of the β-carbonyl to form the C-7 hydroxy
331 group of palmerolide A, is due to module 11, which possesses a KR domain in addition to
332 elongating KS (**Fig. 2**). In the structure of palmerolide A this is followed by the fully reduced
333 subunit from module 12 as discussed above. The final elongation results from module 13, which
334 includes DH and KR domains that contribute to the conjugated ester found as palmerolide A's C-
335 1 through C-3, completing the palmerolide A C_{24} carbon skeleton.
336
337 **Noncanonical termination condensation domain in *palD* for product cyclization and**
338 **release.** Typically, PKS systems terminate with a thioesterase (TE) domain, leading to release of
339 the polyketide from the megaenzyme (71, 81–83). This canonical domain is not present in the *pal*
340 cluster. Instead, the final module in the *cis*-acting biosynthetic gene cluster includes a truncated
341 condensation domain comprised of 133 amino acid residues; compared to the approximately 450
342 residues that comprise a standard condensation domain (84) (**Fig. 2**). Condensation domains
343 catalyze cyclization through ester formation in free-standing condensation domains that act in
344 *trans* as well as in NRPS systems (85, 86). In addition, this non-canonical termination domain is
345 not without precedent in hybrid PKS-NRPS and in PKS systems, as basiliskamide and
346 phormidolide both include condensation domains for product release (53, 87). Though the

347 terminal condensation domain in the *pal* BGC is truncated, it maintains much of the HHXXDDG
348 motif (***SI Appendix*, Fig. S3**), most notably the second histidine, which serves as the catalytic
349 histidine in the condensation reaction (84).
350
351 **Stereochemical and structural confirmation based on sequence information.** KR domains
352 are NADPH-dependent enzymes that belong to the short-chain dehydrogenase superfamily, with
353 Rossman-like folds for co-factor binding (60, 88). Enzymatically, the two KR subtypes, A-Type
354 and B-type, are responsible for stereoselective reduction of β-keto groups and can also
355 determine the stereochemistry of α-substituents. C-type KRs, however, lack reductase activity
356 and often serve as epimerases. A-Type KRs have a key tryptophan residue in the active site, do
357 not possess the LDD amino acid motif, and result in the reduction of β-carbonyls to L-configured
358 hydroxy groups (60). B-Type, which are identified by the presence of an LDD amino acid motif,
359 result in formation of D-configured hydroxy groups (60). The stereochemistry observed in
360 palmerolide A is reflected in the active site sequence information for the D-configured hydroxyl
361 groups from module 5 and module 11 (**Fig. 1** and **Fig. 2**). When an enzymatically active DH
362 domain is within the same module, the stereochemistry of the *cis*- versus *trans*-olefin can be
363 predicted, as the combination of an A-Type KR with a DH results in a *cis*-olefin formation and the
364 combination of a B-Type KR with a DH results in *trans*-olefin formation. The *trans*-α,β-olefins
365 arising from module 7 (Δ^{14}), module 10 (Δ^{8}), and module 13 (Δ^{2}) arise from B-Type KRs and
366 active DHs. The other three olefins present in the structure of palmerolide A, as noted above,
367 likely have positional and stereochemical influence during the enzymatic shifts to the β,γ-
368 positions (Δ^{21} and Δ^{23}) or from the ECH domain (module 6).
369 Additional structural information was obtained through defining the specificity of KS
370 domains using phylogenetic analysis and the *trans*-AT PKS Polyketide Predictor (*trans*ATor)
371 bioinformatic tool (89). KS domains catalyze the sequential two-carbon elongation steps through
372 a Claisen-like condensation with a resulting β-keto feature (90). Additional domains within a given
373 module can modify the β-carbonyl or add functionality to the adjacent α- or γ-positions (60).
374 Specificity of KSs, based the types of modification located on the upstream acetate subunit were
375 determined and were found to be mostly consistent with the retrobiosynthetic predictions (***SI***
376 ***Appendix*, Fig. S4, *SI Appendix*, Table S1**). For example, the first KS, KS1 (module 1), is
377 predicted to receive a subunit containing a β-branch. KS3 (module 4) and KS4 (module 5) are
378 predicted to receive an upstream monomeric unit with α-methylation and an olefinic shift,
379 consistent with the structure of palmerolide A and with the enzymatic transformations resulting
380 from module 3 and module 4, respectively. Interestingly, the KS associated with the HCS
381 cassette is in a clade of its own in the phylogenetic tree. *Trans*ATor also aided in confirming the
382 stereochemical outcomes of the hydroxy groups and olefins, which occur through reduction of the
383 β-carbonyls. The predictions for the D-configured hydroxy groups were consistent with not only
384 the presence of the LDD motif, indicative of type B-type KR as outlined above, but also with
385 stereochemical determination based on the clades of the KS domains of the receiving modules,
386 KS5 (module 6) and KS11 (module 12). They are also consistent with the structure of palmerolide
387 A. The KS predictions, however, did not aid in confirming reduction of the upstream olefins for
388 KS8 (module 9) and KS12 (module 13).
389
390 **Additional *trans*-acting domains and domains between genes responsible for**
391 **biosynthesis.** A glycosyl transferase (*palP*) and lactone oxidase (*palH*) that are associated with
392 glycosylation of polyketides are encoded in the palmerolide A BGC (**Fig. 2**), though glycosylated
393 palmerolides have not been observed. Glycosylation as a means of self-resistance has been
394 described for oleandomycin, which encodes not only two glycosyl transferases, one with
395 substrate specificity toward oleandomycin and the other relaxed specificity, but also a glycosyl
396 hydrolase (91, 92). The glycosylation of the macrolide confers internal resistance to ribosomal
397 inhibition, making the glycosylated macrolide a pro-drug, which is activated by the glycosyl
398 hydrolase that is excreted extracellularly (91, 92). Glycosyl transferases are found in other
399 macrolide- and non-macrolide-producing organisms as a means to inactivate hydroxylated
400 polyketides (93, 94). Though prokaryotic V-ATPases tend to be simpler than those of eukaryotes,

there is homology in the active sites making the pro-drug hypothesis for self-resistance reasonable in palmerolide A biosynthesis (95). The D-arabinono-1,4-lactone oxidase (*palH*) is a FAD-dependent oxidoreductase that likely works in concert with the glycosyltransferase. An ATP-binding cassette (ABC) transporter is encoded between the core biosynthetic genes and the genes for the *trans*-acting enzymes. This transporter, which has homology to SryD and contains the key nucleotide-binding domain GGNGSGKST, may be responsible for the translocation of the macrolide out of the cell. Since it is housed within the BGC, it is likely under the same regulatory control. Additionally, a Band7 protein is encoded. Band7 proteins belong to a ubiquitous family of slipin or stomatin-like integral membrane proteins that are found in all kingdoms, yet lack assignment of function (96). A domain of unknown function (DUF) is annotated; however, this particular DUF, DUF179, has no superfamily relatives in protein homology searches. Together, these *trans*-acting domains decorate the carbon backbone with features that give palmerolide A its unique chemical structure and contribute to its bioactivity.

C. Multiple copies of the *pal* Biosynthetic Gene Cluster Explain Structural Variants in the Palmerolide Family

Careful assembly of the metagenome assembled genome (MAG) of Ca. *S. palmerolidicus* revealed the BGC was present in multiple copies (**Fig. 4** and ***SI Appendix*, Fig. S5-S7**) (44), evidenced by their independent anchoring loci within the MAG and supported by a five-fold increase in depth of coverage relative to the rest of the genome. The structural complexity of the multicopy BGCs represents a biosynthetic system that is similar to that which is found in *Ca*. Didemnitutus mandela, another ascidian-associated verrucomicrobium in the family *Opitutaceae* (71). A total of 5 distinct BGCs consistent with Type-I PKS systems that have much overlap with one another and are likely responsible for much of the structural diversity in the family of palmerolides (17, 18) (**Fig. 4**). Palmerolide A is hypothesized to arise from the BGC designated as *pal* BGC 4 with additional compounds also from this same cluster. The clusters designated as *pal* BGC 1, *pal* BGC 2, *pal* BGC 3, and *pal* BGC 5 have been identified in *Ca.* S. palmerolidicus and the potential products of each are described below. It is hypothesized that there are three levels of diversity introduced to create the family of palmerolides: (1) differences in the site of action for the *trans*-acting domains (with additional *trans*-acting domains at play as well), (2) promiscuity of the initial selection of the starter subunit, and (3) differences in the core biosynthetic genes with additional PKS domains or stereochemical propensities within a module.

There are several palmerolides that likely arise from the same BGC encoding the megaenzyme responsible for palmerolide A (*pal* BGC 4). We hypothesize that the *trans*-acting domains have different sites of action than what is seen in palmerolide A biosynthesis. For example, the chemical scaffold of palmerolide B (**Fig. 4**) is similar to palmerolide A, though the carbamate transfer occurs on the C-7 hydroxy group. Palmerolide B instead bears a sulfate group on the C-11 hydroxy group; proteins with homology to multiple types of sulfatases from the UniProtKB database (P51691, P15289, O69787, Q8ZQJ2) are found in the genome of *Ca.* S. palmerolidicus (44), but are not encoded within the BGCs. One of these *trans*-acting sulfatases likely modifies the molecule post-translationally. Other structural differences including the hydroxylation on C-8 instead of C-10 (as observed in palmerolide A) and the Δ^9 olefin that differs from palmerolide A's Δ^8 olefin, are either due to a difference of the site of action of the LLM (module 9) or a *trans*-acting hydroxylase. Another member of the compound family, palmerolide C, has structural differences attributable to *trans*-acting enzymes as well. Again, a *trans*-acting hydroxylase or the LLM is proposed to be responsible for hydroxylation on C-8. A hydroxy group on C-9 occurs through reduction of the carbonyl. The carbamate installation occurs on C-11 after *trans*-acting hydroxylation or LLM hydroxylation. In addition, the Δ^8 olefin in palmerolide A is not observed, but rather a Δ^6 olefin.

Additional levels of structural variation are seen at the site of the starter unit, likely due to a level of enzymatic promiscuity of the second AT (PalF). This, combined with differences in the sites of action for the *trans*-acting domains, is likely responsible for the structural differences observed in palmerolide F (**Fig. 4**). The terminal olefin on the tail of the macrolide which perhaps

is a product of promiscuity of the selection of the starter unit, the isomeric 3-methyl-3-butenoic acid, is consistent with the aforementioned lack of consensus for malonate selection by the AT. In addition, the KS that receives the starter unit is phylogenetically distinct from the other KS in the *pal* clusters (***SI Appendix*, Fig. S4**).

The retrobiosynthetic hypothesis for palmerolide G (**Fig. 4**) has much similarity to what is present in *pal* BGC 4; however, the presence of a *cis*-olefin rather than a *trans*-olefin could arise from a difference in the enzymatic activity of module 4. This olefin is the result of a shift from an epimerase and, therefore, the stereochemistry is not solely reliant upon the action of the associated KR. Although this difference has not been identified in the BGCs in the samples sequenced, this could be present in other environmental samples that have been batched for processing and compound isolation. Currently, the biosynthetic mechanism is unknown.

The core biosynthetic genes of two palmerolide BGCs (*pal* BGC 1 and *pal* BGC 3) are identical to one another (***SI Appendix*, Fig. S5**) and possess an additional elongation module when compared to *pal* BGC 4. Palmerolide D (**Fig. 4**) is structurally very similar to palmerolide A with the exception of elongation in the carboxylate tail of the macrolide by an isopropyl group. This could arise from one additional round of starter unit elongation via a KS and methylation. These two identical BGCs are consistent with this structural difference. The overall architecture and stereochemistry are otherwise maintained. Palmerolide H (**Fig. 4**) includes the structural differences of both palmerolide B and palmerolide D. It contains the extended carboxylate tail with a terminal olefin and incorporates hydroxylation on C-8 rather than C-10. Again, there is no genomic evidence that this hydroxylation in the α-position is due to incorporation of hydroxymalonate to explain this but is instead likely due to a *trans*-hydroxylase. The carbamate installation occurs on C-7, while sulfonation occurs on C-11 and α-hydroxy placement is on C-8.

The final two palmerolide BGCs both have truncation of the core biosynthetic genes. The gene structure of *pal* BGC 5 (***SI Appendix*, Fig. S6**) shows preservation of genes upstream to the core biosynthetic genes; however, there are no pre-NRP PKS modules noted in the BGC. The HCS cassette, glycosyl transferase, and CT are all present downstream. The predicted product of this cluster does not align with a known palmerolide, though post-translational hydrolysis of the C-24 amide may result in a structure similar to palmerolide E (**Fig. 4**), which maintains much of the structure of palmerolide A; however, is missing the initial polyketide starter unit and the glycine subunit. The final *pal* BGC in *Ca.* S. palmerolidicus, *pal* BGC 2 (***SI Appendix*, Fig. S7**), includes only 5 elongating modules, which would result in a 10-carbon structure that has not been observed. Interestingly, despite the shortened BGC, the HCS cassette, glycosyl transferase, and CT are all present downstream. There would only be a single hydroxy group serving as a substrate for the CT, glycosyl transferase, and sulfatase to act. The 2-carbon site of action for the β-branch introduced in the palmerolide A structures would not be present. The structure-based retrobiosynthesis of the eight known palmerolides (A-F) can be hypothesized to arise from differences in the core biosynthetic genes of these non-identical copies of the *pal* BGC, starter unit promiscuity, and differing sites of action in the *trans*-acting enzymes.

Conclusion

The putative *pal* BGC has been identified and represents the first BGC elucidated from an Antarctic organism. As outlined in a retrobiosynthetic strategy, the *pal* BGC represents a *trans*-AT Type I PKS/NRPS hybrid system with compelling alignment to the predicted biosynthetic steps for palmerolide A. The *pal* BGC is proposed to begin with PKS modules resulting in the incorporation of an isovaleric acid derivative, 3-methylcrotonic acid, as a starter unit, followed by incorporation of a glycine residue with NRPS-type modules. Then, eleven rounds of progressive polyketide elongation likely occur leading to varying degrees of oxidation introduced with each module. There are several interesting non-canonical domains encoded within the BGC, such as an HCS, CT, LLM, and a truncated condensation termination domain. Additionally, a glycosylation domain may be responsible for reversible, pro-drug formation to produce self-resistance to the V-ATPase

activity of palmerolide A. There are additional domains, the function of which have yet to be determined.

A combination of modular alterations, starter unit differences, and activity of *trans*-acting enzymes contributes to Nature's production of a suite of palmerolide analogues. There are a total of five distinct *pal* BGCs in the MAG of *Ca.* S. palmerolidicus, yielding the known eight palmerolides, with genetic differences that explain some of the structural variety seen within this family of compounds. These include differences in modules that comprise the core biosynthetic genes. Additionally, it is proposed that some of the architectural diversity of palmerolides arises from different sites of action of the *trans*-acting, or non-colinear, modules. Starter unit promiscuity is another potential source of the structural differences observed in the compounds. Analysis of the *pal* BGC not only provides insight into the architecture of this Type I PKS/NRPS hybrid BGC with unique features, but also lays the foundational groundwork for drug development studies of palmerolide A via heterologous expression.

Materials and Methods

The details of field collections, sample processing and genomic methods (sequencing, assembly, and analysis) used to identify the putative biosynthetic gene clusters are described in detail in two related publications (35, 44). In brief, *Synoicum adareanum* samples were collected by SCUBA from the Antarctic Peninsula in the Anvers Island Archipelago and flash frozen. Microbial cells were separated from host tissue using a homogenization protocol established for this holobiont followed by differential centrifugation to separate the host cellular debris and microbial cells (31). The metagenome assembled genome (MAG) sequence analyzed in this study was generated from two *S. adareanum* samples (Bon-1c-2011 and Del-2b-2011) with high copy numbers of the putative biosynthetic gene cluster sequenced using PacBio technology. The resulting assembly of produced a nearly complete 4.3 MB genome, with five unique contigs and five varying copies of the *pal* BGC (referred to as *pal* BGC 1 through *pal* BGC 5) (44). Given the phylogenomic novelty of the MAG, the name *Candidatus* Synoicihabitans palmerolidicus was proposed for this new genus in the *Opitutaceae* family (Verrucomicrobia phylum).

The methods employed in this study used bioinformatic tools to develop predictive models of palmerolide biosynthesis. Enzymatic reactions and organic synthetic interpretations were based on homology analyses. Automated annotation and manual bioinformatic tools were used to discern the details of palmerolide A biosynthesis in addition to generating predictions for the other BGCs. The *Ca*. S. palmerolidicus MAG was annotated using antiSMASH (v. 5.0) (51) using the full complement of annotation options available. Then we predicted the gene cluster responsible for palmerolide A biosynthesis using retrobiosynthetic predictions focused on the 5' end of the BGCs (Fig. 1). Only one of the five BGCs met the criteria for a non-ribosomal peptide glycine starter unit. The annotation predictions were integrated and validated with results of additional protein family homology analysis, conserved domain searches, active site and motif identification to predict the step-wise biosynthesis of palmerolide A. Manual annotation of the *pal* BGC sequences included BLASTP searches to confirm enzymatic identities, then protein family alignments were used to identify active site residues key for stereochemical outcomes, confirm substrate affinities, and other biochemical synthesis details.

Additional manual bioinformatic efforts included obtaining BGCs from public NCBI databases for basiliskamide, bryostatin 1, calyculin, corallopyronin, mandelalide, onnamide, oxazolamycin, pederin, phormidolide, psymberin, sorangicin, and myxoviricin (***SI Appendix*, Table S2**). ClustalO alignment tool in the CLC Genome Workbench (QIAGEN aarhau A/S v. 20.0.3) was used for multiple sequence alignments of enzymatic domains with HMM Pfam Seeds obtained from EMBL-EBI and the amino acid sequences from the other PKS BGCs. MIBiG (97) was used to acquire the KS amino acid sequence from the type III PKS BGC responsible for 3-(2'-hydroxy-3'-oxo-4'-methylpentyl)-indole biosynthesis from *Xenorhabdus bovienii SS-2004* (GenBank Accession: FN667741.1), which was used for an outgroup. The *pal* BGC ACPs and

PCPs were numbered according to their position in the proposed biosynthesis of palmerolide A (***SI Appendix*, Fig. S1 and S2**). The BGC KSs were numbered according to their position in their proposed biosynthesis in the literature. Prior to the construction of the phylogenetic tree for the KS domains (***SI Appendix*, Fig. S4**), the sequences in the alignment were manually inspected and trimmed. Phylogenetic trees were created in the CLC Genome Workbench (QIAGEN aarhau A/S v. 20.0.3) with Neighbour Joining (NJ) as a distance method and Bayesian estimation for ACP and PCP comparisons as well as for KS analysis. Jukes-Cantor was selected for the genetic distance model and bootstrapping with performed with 100 replicates. Additionally, the sequence of each KS in the *pal* BGCs was queried using the *trans*-AT PKS Polyketide Predictor (*trans*ATor) to help define the specificity of KS domains. The software is based on phylogenetic analyses of fifty-four *trans*-AT type I PKS systems with 655 KS sequences and the resulting clades are referenced to help predict the KS specificity for the upstream unit (89).

Data Deposition

Data from this project has been deposited to the NCBI under BioProject Accession Number PRJNA662631. The GenBank Accession Number for the MAG of *Ca.* Synoicohabitans palmerolidicus is JAGGDC000000000. The BGCs can be accessed in the MiBIG database with the associated Accession Numbers: BGC0002118 (for *pal* BGC 4) and BGC0002119 (for *pal* BGC 3).

Acknowledgments

Support for this research was provided in part by the National Institutes of Health award (CA205932) to A.E.M., B.J.B., and P.S.G.C., with additional support from National Science Foundation awards (ANT-0838776, and PLR-1341339 to B.J.B., ANT-0632389 to A.E.M.) The authors acknowledge the assistance of field team members, including William Dent, Charles D. Amsler, James B. McClintock, Margaret O. Amsler, and Katherine Schoenrock. This work would not have been possible without the outstanding logistical support of the United States Antarctic Program. Lucas Bishop, Robert Read, and Mary L. Higham are also recognized for their contributions.

References

1. A. R. Carroll, B. R. Copp, R. A. Davis, R. A. Keyzers, M. R. Prinsep, Marine natural products. *Nat. Prod. Rep.* **36**, 122–173 (2019).
2. C. A. Dejong, *et al.*, Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. *Nat. Chem. Biol.* **12**, 1007–1014 (2016).
3. J. G. Lawrence, H. Ochman, M. A. Ragan, Reconciling the many faces of lateral gene transfer. *Trends Microbiol.* **10**, 1–4 (2002).
4. M. Ravenhall, N. Škunca, F. Lassalle, C. Dessimoz, Inferring Horizontal Gene Transfer. *PLoS Comput. Biol.* **11**, 1–16 (2015).
5. I. Schmitt, H. T. Lumbsch, Ancient horizontal gene transfer from bacteria enhances biosynthetic capabilities of fungi. *PLoS One* **4**, 1–8 (2009).
6. E. W. Schmidt, Trading molecules and tracking targets in symbiotic interactions. *Nat. Chem. Biol.* **4**, 466–473 (2008).
7. E. W. Schmidt, The secret to a successful relationship: Lasting chemistry between ascidians and their symbiotic bacteria. *Invertebr. Biol.* **134**, 88–102 (2015).
8. E. J. N. Helfrich, J. Piel, Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat. Prod. Rep.* **33**, 231–316 (2016).
9. A. M. P. Koskinen, K. Karisalmi, Polyketide stereotetrads in natural products. *Chem. Soc. Rev.* **34**, 677–690 (2005).

616 10. J. Napolitano, A. Daranas, M. Norte, J. Fernandez, Marine Macrolides, a Promising
617 Source of Antitumor Compounds. *Anticancer. Agents Med. Chem.* **9**, 122–137 (2012).
618 11. R. Ueoka, *et al.*, Metabolic and evolutionary origin of actin-binding polyketides from
619 diverse organisms. *Nat. Chem. Biol.* **11**, 705–712 (2015).
620 12. M. E. Bordeleau, *et al.*, Stimulation of mammalian translation initiation factor eIF4A activity
621 by a small molecule inhibitor of eukaryotic translation. *Proc. Natl. Acad. Sci. U. S. A.* **102**,
622 10460–10465 (2005).
623 13. S. Nishimura, *et al.*, Structure-activity relationship study on 13-deoxytedanolide, a highly
624 antitumor macrolide from the marine sponge *Mycale adhaerens*. *Bioorganic Med. Chem.*
625 **13**, 455–462 (2005).
626 14. R. Shen, C. T. Lin, E. J. Bowman, B. J. Bowman, J. A. Porco, Lobatamide C: Total
627 synthesis, stereochemical assignment, preparation of simplified analogues, and V-ATPase
628 inhibition studies. *J. Am. Chem. Soc.* **125**, 7889–7901 (2003).
629 15. E. J. Bowman, K. R. Gustafson, B. J. Bowman, M. R. Boyd, Identification of a New
630 Chondropsin Class of Antitumor Compound that Selectively Inhibits V-ATPases. *J. Biol.*
631 *Chem.* **278**, 44147–44152 (2003).
632 16. S. Kazami, *et al.*, Iejimalides show anti-osteoclast activity via V-ATPase inhibition. *Biosci.*
633 *Biotechnol. Biochem.* **70**, 1364–1370 (2006).
634 17. T. Diyabalanage, C. D. Amsler, J. B. McClintock, B. J. Baker, Palmerolide A, a cytotoxic
635 macrolide from the antarctic tunicate *Synoicum adareanum*. *J. Am. Chem. Soc.* **128**,
636 5630–5631 (2006).
637 18. J. H. Noguez, *et al.*, Palmerolide macrolides from the Antarctic tunicate *Synoicum*
638 *adareanum*. *Bioorganic Med. Chem.* **19**, 6608–6614 (2011).
639 19. B. Shen, *et al.*, Polyketide biosynthesis beyond the type I, II, and III polyketide synthase
640 paradigms: A progress report. *ACS Symp. Ser.* **955**, 154–166 (2007).
641 20. B. Wilkinson, *et al.*, Novel octaketide macrolides related to 6-deoxyerythronolide B provide
642 evidence for iterative operation of the erythromycin polyketide synthase. *Chem. Biol.* **7**,
643 111–117 (2000).
644 21. S. Tatsuno, K. Arakawa, H. Kinashi, Analysis of Modular-iterative Mixed Biosynthesis of
645 Lankacidin by Heterologous Expression and Gene Fusion. *J. Antibiot. (Tokyo).* **60**, 700–
646 708 (2007).
647 22. B. Wang, F. Guo, C. Huang, H. Zhao, Unraveling the iterative type I polyketide synthases
648 hidden in *Streptomyces*. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 8449–8454 (2020).
649 23. T. A. Nguyen, *et al.*, Exploiting the mosaic structure of *trans*-acyltransferase polyketide
650 synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* **26**, 225–
651 233 (2008).
652 24. S. F. Haydock, *et al.*, Divergent sequence motifs correlated with the substrate specificity of
653 (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide
654 synthases. *FEBS Lett.* **374**, 246–248 (1995).
655 25. H. Jenke-Kodama, A. Sandmann, R. Müller, E. Dittmann, Evolutionary implications of
656 bacterial polyketide synthases. *Mol. Biol. Evol.* **22**, 2027–2039 (2005).
657 26. A. Nivina, K. P. Yuet, J. Hsu, C. Khosla, Evolution and Diversity of Assembly-Line
658 Polyketide Synthases. *Chem. Rev.* **119**, 12524–12547 (2019).
659 27. P. Videau, K. N. Wells, A. J. Singh, W. H. Gerwick, B. Philmus, Assessment of *Anabaena*
660 sp. Strain PCC 7120 as a Heterologous Expression Host for Cyanobacterial Natural
661 Products: Production of Lyngbyatoxin A. *ACS Synth. Biol.* **5**, 978–988 (2016).
662 28. E. J. Kim, *et al.*, Heterologous Production of 4-O-Demethylbarbamide, a Marine
663 Cyanobacterial Natural Product. *Org. Lett.* **14**, 5824–5827 (2012).
664 29. C. Greunke, *et al.*, Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic
665 potential. *Metab. Eng.* **47**, 334–345 (2018).
666 30. J. B. McClintock, C. D. Amsler, B. J. Baker, R. W. M. Van Soest, Ecology of antarctic
667 marine sponges: An overview. *Integr. Comp. Biol.* **45**, 359–368 (2005).
668 31. A. Clarke, J. A. Crame, Evolutionary dynamics at high latitudes: Speciation and extinction
669 in polar marine faunas. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 3655–3666 (2010).

32. R. M. Young, *et al.*, Site-specific variability in the chemical diversity of the Antarctic red alga *Plocamium cartilagineum*. *Mar. Drugs* **11**, 2126–2139 (2013).
33. K. D. Paull, E. Hamel, L. Malspeis, Prediction of Biochemical Mechanism of Action from the In Vitro Antitumor Screen of the National Cancer Institute. *Cancer Chemother. Agents*, 9–45 (1995).
34. K. Von Schwarzenberg, *et al.*, Mode of cell death induction by pharmacological vacuolar H^+-ATPase (V-ATPase) inhibition. *J. Biol. Chem.* **288**, 1385–1396 (2013).
35. A. E. Murray, *et al.*, Uncovering the Core Microbiome and Distribution of Palmerolide in *Synoicum adareanum* Across the Anvers Island Archipelago, Antarctica. *Mar. Drugs* **18**, 298 (2020).
36. K. P. Kaliappan, P. Gowrisankar, Synthetic studies on a marine natural product, palmerolide A: Synthesis of C1-C9 and C15-C21 fragments. *Synlett*, 1537–1540 (2007).
37. M. P. Lisboa, D. M. Jones, G. B. Dudley, Formal synthesis of palmerolide A, featuring alkynogenic fragmentation and syn-selective vinylogous aldol chemistry. *Org. Lett.* **15**, 886–889 (2013).
38. M. D. Lebar, B. J. Baker, Synthesis of the C3-14 fragment of palmerolide A using a chiral pool based strategy. *Tetrahedron* **66**, 1557–1562 (2010).
39. Z. K. Wen, Y. H. Xu, T. P. Loh, Palladium-catalyzed cross-coupling of unactivated alkenes with acrylates: Application to the synthesis of the C13-C21 fragment of palmerolide A. *Chem. - A Eur. J.* **18**, 13284–13287 (2012).
40. S. A. Pujari, P. Gowrisankar, K. P. Kaliappan, A shimizu non-aldol approach to the formal total synthesis of palmerolide A. *Chem. - An Asian J.* **6**, 3137–3151 (2011).
41. K. R. Prasad, A. B. Pawar, Stereoselective synthesis of C1-C18 region of palmerolide A from tartaric acid. *Synlett* **4**, 1093–1095 (2010).
42. A. B. Pawar, K. R. Prasad, Formal total synthesis of palmerolide A. *Chem. - A Eur. J.* **18**, 15202–15206 (2012).
43. C. S. Riesenfeld, A. E. Murray, B. J. Baker, Characterization of the microbial community and polyketide biosynthetic potential in the palmerolide-producing tunicate *Synoicum adareanum*. *J. Nat. Prod.* **71**, 1812–1818 (2008).
44. A. E. Murray, *et al.*, Discovery of an Antarctic ascidian-associated uncultivated Verrucomicrobia that encodes antimelanoma palmerolide biosynthetic capacity. *BioRxiv*. (Preprint: 10.1101/2021.05.05.442870) (2021).
45. W. Mattheus, *et al.*, Isolation and Purification of a New Kalimantacin/Batumin-Related Polyketide Antibiotic and Elucidation of Its Biosynthesis Gene Cluster. *Chem. Biol.* **17**, 149–159 (2010).
46. S. Sudek, *et al.*, Identification of the putative bryostatin polyketide synthase gene cluster from "*Candidatus* Endobugula sertula", the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. *J. Nat. Prod.* **70**, 67–74 (2007).
47. D. J. Edwards, *et al.*, Structure and Biosynthesis of the Jamaicamides, New Mixed Polyketide-Peptide Neurotoxins from the Marine Cyanobacterium *Lyngbya majuscula*. *Chem. Biol.* **11**, 817–833 (2004).
48. Z. Chang, *et al.*, Biosynthetic Pathway and Gene Cluster Analysis of Curacin A, an Antitubulin Natural Product from the Tropical Marine Cyanobacterium *Lyngbya majuscula* †. *J. Nat. Prod.* **67**, 1356–1367 (2004).
49. M. A. Matilla, H. Stöckmann, F. J. Leeper, G. P. C. Salmond, Bacterial biosynthetic gene clusters encoding the anti-cancer haterumalide class of molecules: Biogenesis of the broad spectrum antifungal and anti-oomycete compound, oocydin A. *J. Biol. Chem.* **287**, 39125–39138 (2012).
50. K. M. Fisch, *et al.*, Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat. Chem. Biol.* **5**, 494–501 (2009).
51. K. Blin, *et al.*, antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, W81–W87 (2019).
52. T. A. M. Gulder, M. F. Freeman, J. Piel, The Catalytic Diversity of Multimodular Polyketide Synthases: Natural Product Biosynthesis Beyond Textbook Assembly Rules (2011)

724 https:/doi.org/10.1007/128_2010_113.
725 53. M. J. Bertin, *et al.*, The Phormidolide Biosynthetic Gene Cluster: A *trans*-AT PKS Pathway
726 Encoding a Toxic Macrocyclic Polyketide. *ChemBioChem* **17**, 164–173 (2016).
727 54. C. D. Reeves, *et al.*, Alteration of the substrate specificity of a modular polyketide
728 synthase acyltransferase domain through site-specific mutations. *Biochemistry* **40**, 15464–
729 15470 (2001).
730 55. G. Yadav, R. S. Gokhale, D. Mohanty, Computational approach for prediction of domain
731 organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.* **328**,
732 335–363 (2003).
733 56. B. J. Dunn, K. R. Watts, T. Robbins, D. E. Cane, C. Khosla, Comparative analysis of the
734 substrate specificity of *trans*- versus *cis*-acyltransferases of assembly line polyketide
735 synthases. *Biochemistry* **53**, 3796–3806 (2014).
736 57. A. L. Díaz-Pérez, C. Díaz-Pérez, J. Campos-García, Bacterial l-leucine catabolism as a
737 source of secondary metabolites. *Rev. Environ. Sci. Bio/Technology* **15**, 1–29 (2016).
738 58. T. Mahmud, *et al.*, A novel biosynthetic pathway providing precursors for fatty acid
739 biosynthesis and secondary metabolite formation in myxobacteria. *J. Biol. Chem.* **277**,
740 32768–32774 (2002).
741 59. M. Röttig, *et al.*, NRPSpredictor2 - A web server for predicting NRPS adenylation domain
742 specificity. *Nucleic Acids Res.* **39**, 362–367 (2011).
743 60. A. T. Keatinge-Clay, The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29**,
744 1050 (2012).
745 61. Ö. Erol, *et al.*, Biosynthesis of the myxobacterial antibiotic corallopyronin A.
746 *ChemBioChem* **11**, 1253–1265 (2010).
747 62. C. Zhao, *et al.*, Oxazolomycin biosynthesis in *Streptomyces albus* JA3453 featuring an
748 “acyltransferase-less” type I polyketide synthase that incorporates two distinct extender
749 units. *J. Biol. Chem.* **285**, 20097–20108 (2010).
750 63. H. Irschik, *et al.*, Analysis of the sorangicin gene cluster reinforces the utility of a combined
751 phylogenetic/retrobiosynthetic analysis for deciphering natural product assembly by *trans*-
752 AT PKS. *ChemBioChem* **11**, 1840–1849 (2010).
753 64. D. C. Cantu, Y. Chen, P. J. Reilly, Thioesterases: A new perspective based on their
754 primary and tertiary structures. *Protein Sci.* **19**, 1281–1295 (2010).
755 65. J. Moldenhauer, *et al.*, The Final Steps of Bacillaene Biosynthesis in *Bacillus*
756 *amyloliquefaciens* FZB42: Direct Evidence for β,γ Dehydration by a *trans*-Acyltransferase
757 Polyketide Synthase. *Angew. Chemie* **122**, 1507–1509 (2010).
758 66. G. Berkhan, C. Merten, C. Holec, F. Hahn, The Interplay between a Multifunctional
759 Dehydratase Domain and a C-Methyltransferase Effects Olefin Shift in Ambruticin
760 Biosynthesis. *Angew. Chemie - Int. Ed.* **55**, 13589–13592 (2016).
761 67. C. Hertweck, The biosynthetic logic of polyketide diversity. *Angew. Chemie - Int. Ed.* **48**,
762 4688–4716 (2009).
763 68. J. Moldenhauer, X. H. Chen, R. Borriss, J. Piel, Biosynthesis of the antibiotic bacillaene,
764 the product of a giant polyketide synthase complex of the *trans*-AT family. *Angew. Chemie*
765 *- Int. Ed.* **46**, 8195–8197 (2007).
766 69. T. Wakimoto, *et al.*, Calyculin biogenesis from a pyrophosphate protoxin produced by a
767 sponge symbiont. *Nat. Chem. Biol.* **10**, 648–655 (2014).
768 70. E. Esquenazi, *et al.*, Visualizing the spatial distribution of secondary metabolites produced
769 by marine cyanobacteria and sponges via MALDI-TOF imaging. *Mol. Biosyst.* **4**, 562–570
770 (2008).
771 71. J. Lopera, I. J. Miller, K. L. McPhail, J. C. Kwan, Increased Biosynthetic Gene Dosage in a
772 Genome-Reduced Defensive Bacterial Symbiont. *mSystems* **2** (2017).
773 72. T. J. Buchholz, *et al.*, Polyketide β-Branching in Bryostatin Biosynthesis: Identification of
774 Surrogate Acetyl-ACP Donors for BryR, an HMG-ACP Synthase. *Chem. Biol.* **17**, 1092–
775 1100 (2010).
776 73. Y. Q. Cheng, G. L. Tang, B. Shen, Type I polyketide synthase requiring a discrete
777 acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3149–

3154 (2003).
74. K. C. Nicolaou, *et al.*, Chemical synthesis and biological evaluation of palmerolide A analogues. *J. Am. Chem. Soc.* **130**, 10019–10023 (2008).
75. S. F. Haydock, *et al.*, Organization of the biosynthetic gene cluster for the macrolide concanamycin A in *Streptomyces neyagawaensis* ATCC 27449. *Microbiology* **151**, 3161–3169 (2005).
76. W. Chen, *et al.*, Characterization of the Polyoxin Biosynthetic Gene Cluster from *Streptomyces cacaoi* and Engineered Production of Polyoxin H. *J. Biol. Chem.* **284**, 10627–10638 (2009).
77. T. K. Mihali, W. W. Carmichael, B. A. Neilan, A putative gene cluster from a *Lyngbya wollei* bloom that encodes paralytic shellfish toxin biosynthesis. *PLoS One* **6** (2011).
78. A. K. El-Sayed, J. Hothersall, C. M. Thomas, Quorum-sensing-dependent regulation of biosynthesis of the polyketide antiobiotic mupirocin in *Pseudomonas fluorescens* NCIMB 10586. *Microbiology* **147**, 2127–2139 (2001).
79. S. Maier, *et al.*, Functional characterization of different ORFs including luciferase-like monooxygenase genes from the mensacarcin gene cluster. *ChemBioChem* **16**, 1175–1182 (2015).
80. G. Lackner, E. E. Peters, E. J. N. Helfrich, J. Piel, Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. *Proc. Natl. Acad. Sci.* **114**, E347–E356 (2017).
81. J. J. Gehret, *et al.*, Terminal alkene formation by the thioesterase of curacin A biosynthesis: Structure of a decarboxylating thioesterase. *J. Biol. Chem.* **286**, 14445–14454 (2011).
82. L. Gu, *et al.*, Metamorphic enzyme assembly in polyketide diversification. *Nature* **459**, 731–735 (2009).
83. J. Piel, A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14002–14007 (2002).
84. T. Stachelhaus, H. D. Mootz, V. Bergendahl, M. A. Marahiel, Peptide Bond Formation in Nonribosomal Peptide Biosynthesis. *J. Biol. Chem.* **273**, 22773–22781 (1998).
85. S. Lin, S. G. Van Lanen, B. Shen, A free-standing condensation enzyme catalyzing ester bond formation in C-1027 biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 4183–4188 (2009).
86. K. Zaleta-Rivera, *et al.*, A bidomain nonribosomal peptide synthetase encoded by FUM14 catalyzes the formation of tricarballylic esters in the biosynthesis of fumonisins. *Biochemistry* **45**, 2561–2569 (2006).
87. C. M. Theodore, *et al.*, Genomic and metabolomic insights into the natural product biosynthetic diversity of a feral-hog-associated *Brevibacillus laterosporus* strain. *PLoS One* **9**, 3–12 (2014).
88. A. T. Keatinge-Clay, R. M. Stroud, The Structure of a Ketoreductase Determines the Organization of the β-Carbon Processing Enzymes of Modular Polyketide Synthases. *Structure* **14**, 737–748 (2006).
89. E. J. N. Helfrich, *et al.*, Automated structure prediction of *trans*-acyltransferase polyketide synthase products. *Nat. Chem. Biol.* **15**, 813–821 (2019).
90. C. Khosla, Y. Tang, A. Y. Chen, N. A. Schnarr, D. E. Cane, Structure and Mechanism of the 6-Deoxyerythronolide B Synthase. *Annu. Rev. Biochem.* **76**, 195–221 (2007).
91. L. M. Quirós, I. Aguirrezabalaga, C. Olano, C. Méndez, J. A. Salas, Two glycosyltransferases and a glycosidase are involved in oleandomycin modification during its biosynthesis by *Streptomyces antibioticus*. *Mol. Microbiol.* **28**, 1177–1185 (1998).
92. T. A. Wencewicz, Crossroads of Antibiotic Resistance and Biosynthesis. *J. Mol. Biol.* **431**, 3370–3399 (2019).
93. G. Jenkins, E. Cundliffe, Cloning and characterization of two genes from *Streptomyces lividans* that confer inducible resistance to lincomycin and macrolide antibiotics. *Gene* **108**, 55–62 (1991).

94. A. Gourmelen, M. H. Blondelet-Rouault, J. L. Pernodet, Characterization of a glycosyl transferase inactivating macrolides, encoded by gimA from *Streptomyces ambofaciens*. *Antimicrob. Agents Chemother.* **42**, 2612–2619 (1998).
95. K. Yokoyama, H. Imamura, Rotation, structure, and classification of prokaryotic V-ATPase. *J. Bioenerg. Biomembr.* **37**, 405–410 (2005).
96. M. Boehm, *et al.*, Structural and mutational analysis of band 7 proteins in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* **191**, 6425–6435 (2009).
97. S. A. Kautsar, *et al.*, MIBiG 2.0: A repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* **48**, D454–D458 (2020).

Figures

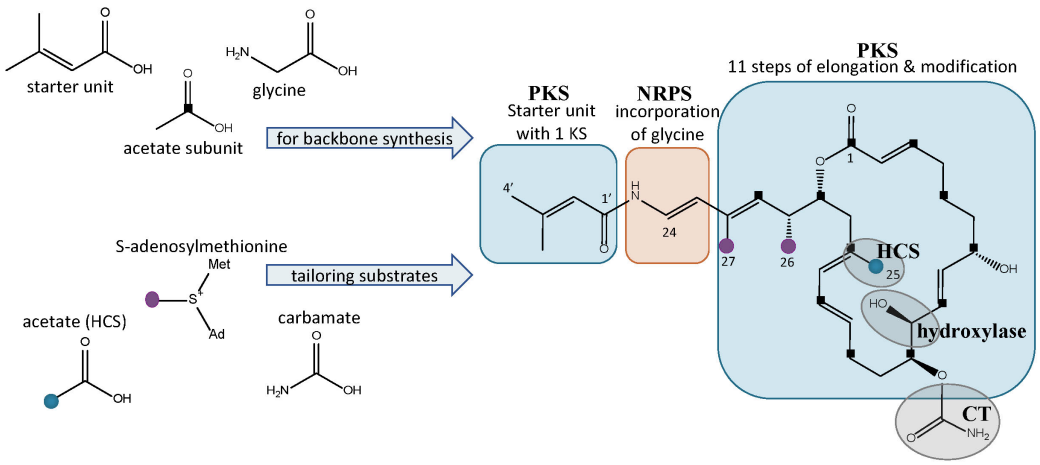


Fig. 1. Structure of palmerolide A with notations for the proposed retrobiosynthesis. Backbone synthesis is a result of incorporation of the starter unit, a glycine residue, and acetate subunits (C1 indicated by black squares). Structural features from *trans*-acting tailoring enzymes (indicated by grey ovals) utilize additional substrates: methyl transfers from SAM (purple dots), installation of C-25 methyl from acetate (blue dot) via an HCS cassette, and carbamoyl transfer to the secondary alcohol on C-11. The α-hydroxy group on C-10 is predicted to arise from incorporation of hydroxymalonic acid or a *trans*-acting hydroxylase.

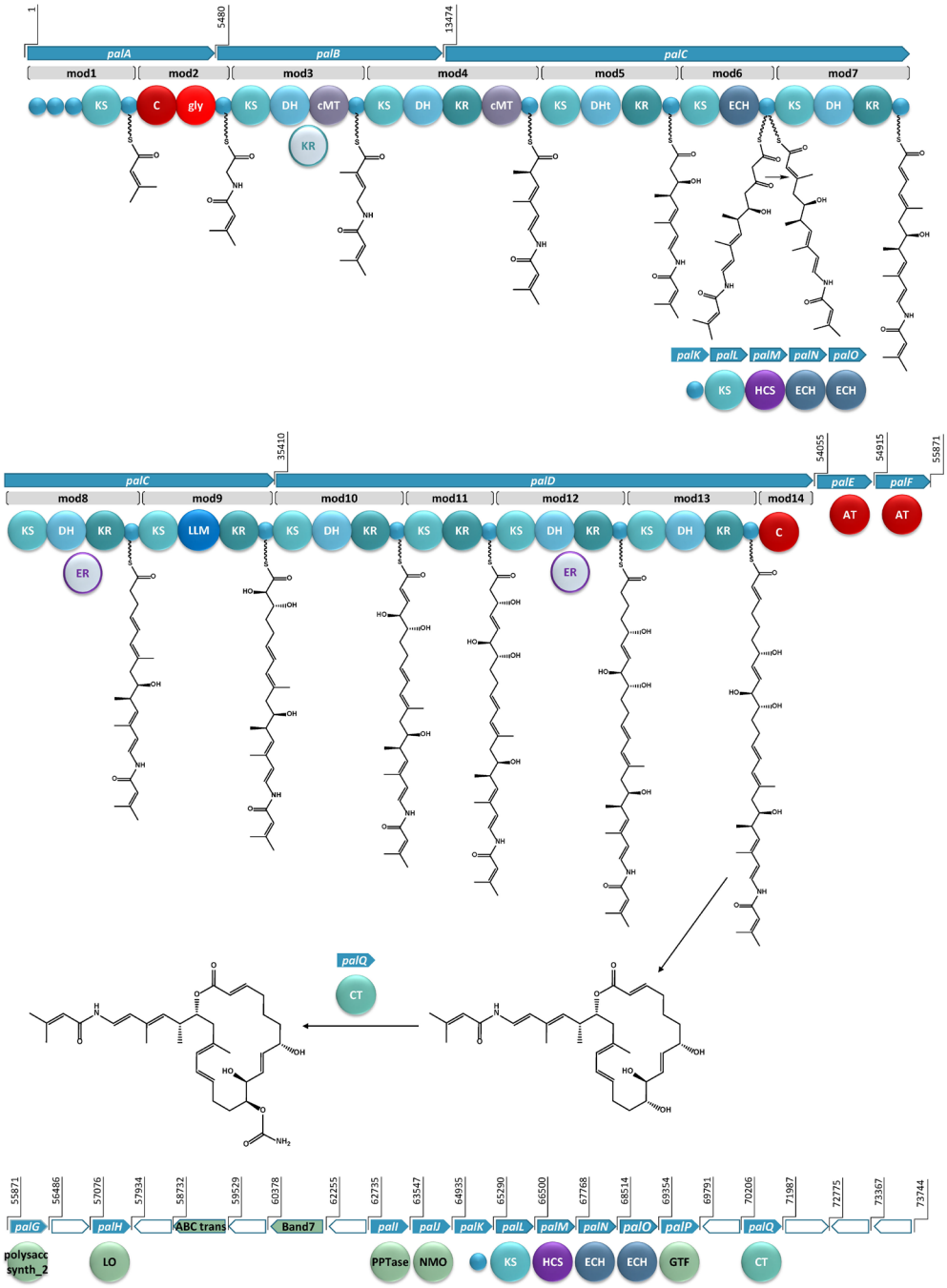
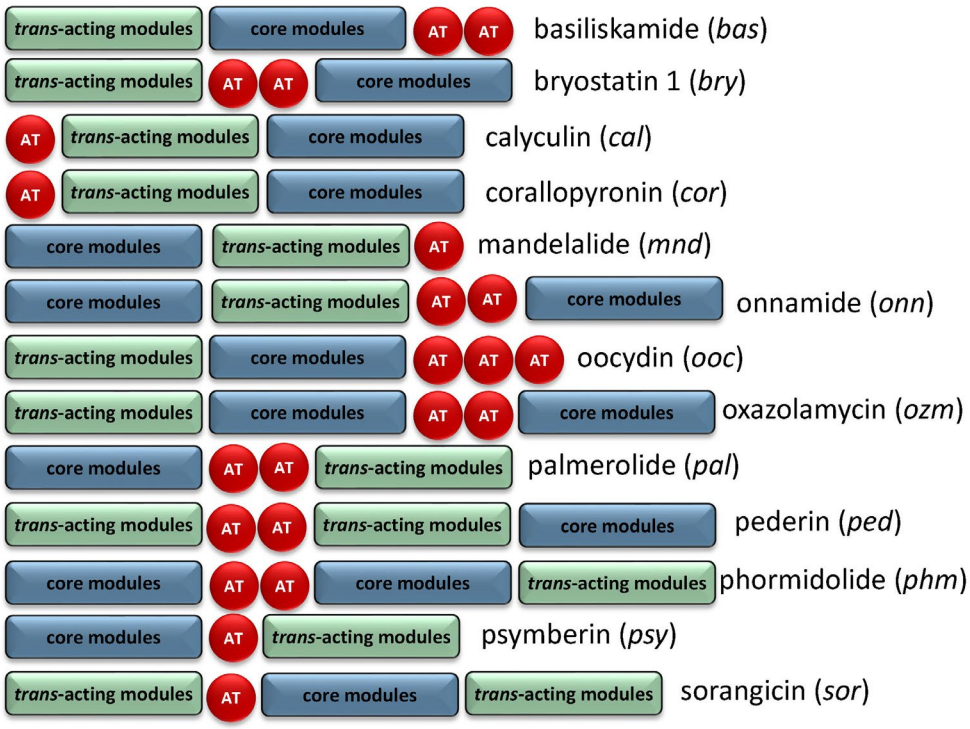


Fig. 2. The proposed BGC for palmerolide A, showing the hybrid PKS-NRPS system. KS: ketosynthase domain, C: condensation domain, gly: adenylation domain for glycine incorporation, DH: dehydratase domain, cMT: carbon methyl transferase domain, KR: ketoreductase domain, DHt: dehydratase variant; ECH: enoyl-CoA hydratase, LLM: luciferase-like monooxygenase, AT: acyl transferase; polysacc synt_2: polysaccharide biosynthesis protein, LO: lactone oxidase, ABC

863 trans: ATP-binding cassette transporter, Band7: stomatin-like integral membrane, PPTase:
864 phosphopantetheinyl transferase, NMO: nitronate monooxygenase, HCS: hydroxymethylglutaryl-
865 CoA synthase, GTF: glycosyl transferase ER: enoyl reductase, CT: carbamoyl transferase, small
866 blue circles represent acyl- or peptidyl-carrier proteins. Ppant arms are symbolized by wavy lines.
867 The grey domains indicate domains that would be expected to perform an enzymatic
868 transformation; however, are not encoded in the BGC. Blue arrows indicate biosynthetic genes.
869 Green arrows indicate genes that encode for non-biosynthetic proteins. White arrows reflect
870 hypothetical genes. The BGC is displayed in reverse compliment.
871

872



873
874 **Fig. 3.** Comparison of BGC organization of select *trans*-AT systems. There is significant
875 variability in the order of the core modules, AT modules, and modules which contain *trans*-acting
876 tailoring enzymes. There is also variability in the number of encoded AT modules, though the AT
877 modules are typically encoded on separate, but tandem genes if more than one is present.
878
879

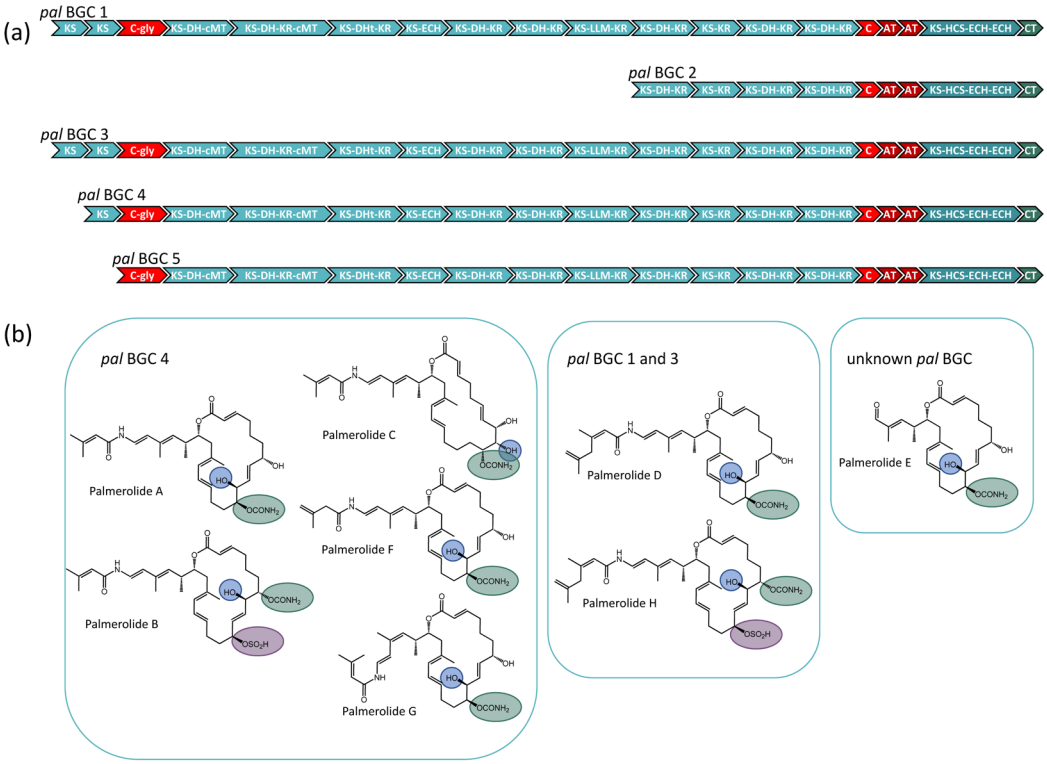


Fig. 4 (a) Comparison of the modular structure of the 5 *pal* BGCs. **(b)** Family of palmerolides. Much of the structural diversity can be explained by differences due to starter unit promiscuity, sites of action for the *trans*-acting tailoring enzymes, and differences in the core modules of the multiple *pal* BGCs. It is proposed that *pal* BGC 4 is responsible for not only palmerolide A, but also palmerolide B, palmerolide C, palmerolide F, and palmerolide G. It is interesting to note that the modular structure of the domains responsible for biosynthesis are equivalent for *pal* BGC 1 and *pal* BGC 3. These two BGCs contain an additional KS domain as compared to *pal* BGC 4 and are likely responsible for the biosynthesis of palmerolide D and palmerolide H.