

Animal FAS-like polyketide synthases produce diverse polypropionates

Feng Li^a, Zhenjian Lin^a, Patrick J Krug^b, J. Leon Catrow^c, James E. Cox^{c,d} and Eric W. Schmidt^a

^aDepartment of Medicinal Chemistry, University of Utah, Salt Lake City, UT, 84112, USA;

^bDepartment of Biological Sciences, California State University, Los Angeles, CA, 90032, USA;

^cMetabolomics Core, Health Sciences Center, Salt Lake City, Utah 84112, USA;

^dDepartment of Biochemistry, University of Utah, Salt Lake City, Utah 84112, USA.

* Eric W. Schmidt.

Email: ews1@utah.edu

Author Contributions: The manuscript was written through contributions of all authors. E.S. and F.L. designed research; Z.L., P.K., and F.L. performed research; J.L.C. and J.E.C. performed MS² analysis; F.L., Z.L., E.S., and P.K. wrote the paper.

Competing Interest Statement: All authors declare no interest conflict.

Classification: Biological sciences; Biochemistry.

Keywords: metazoan biosynthesis; sacoglossan mollusc; type I PKS

This PDF file includes:

Main Text

Abstract

Animal cytoplasmic fatty acid synthase (FAS) represents a unique family of enzymes that are classically thought to be most closely related to fungal polyketide synthase (PKS). Recently, a widespread new family of animal lipid metabolic enzymes has been described that bridges the gap between these two ubiquitous and important enzyme classes: the animal FAS-like PKSs (AFPks). Although very similar in sequence to FAS enzymes that produce saturated lipids widely found in animals, AFPks instead produce structurally diverse compounds that resemble bioactive polyketides. Little is known about the factors that bridge lipid and polyketide synthesis in the animals. Here, we describe the function of EcPKS2 from *Elysia chlorotica*, which synthesizes a complex polypropionate natural product found in this mollusc. EcPKS2 starter unit promiscuity potentially explains the high diversity of polyketides found in and among molluscan species. Biochemical comparison of EcPKS2 with the previously described EcPKS1 reveals molecular principles governing substrate selectivity that should apply to related enzymes encoded within the genomes of photosynthetic gastropods. Hybridization experiments combining EcPKS1 and EcPKS2 demonstrate the interactions between the ketoreductase and ketosynthase domains in governing the product outcomes. Overall, these findings enable an understanding of the molecular principles of structural diversity underlying the many molluscan polyketides likely produced by the diverse and novel AFPK enzyme family.

Significance Statement

The animal FAS-like PKSs (AFPks) are phylogenetically widespread enzymes, bridging the evolutionary gap between lipids and bioactive compounds known as polyketides in animals. Here, we demonstrate key functions of this enzyme class that defy expectation for how fatty acid or polyketide biosynthesis should work, revealing an untapped and previously unknown source of chemical richness in the animals. These results explain how different proteins encoded in the animal genome contribute to the diversity of ecologically important compounds in animals, and suggest a tantalizing wealth of unexpected enzymes and important chemicals remains to be discovered.

Main Text

Introduction

Much of lipid metabolism is founded in a ubiquitous family of related enzymes, the fatty acid synthases (FASs) and polyketide synthases (PKSs). In most animals including humans, a unique type I FAS is found in the cytoplasm, where it synthesizes essential saturated fatty acids (1). The current view of animal FAS (aFAS) is that it originated from a type I modular PKS, very similar to the highly-reducing (HR) PKSs found in fungi. Indeed, the phylogenetic tree of PKS and animal FAS enzymes reveals that related PKSs from animals, fungi, and bacteria, identified in a single clade, while the aFASs form a distinctly separate clade (Fig. 1). Recently, we described a widespread new family, dubbed the aFAS-like PKSs (AFPks), that are more similar to aFAS than to PKSs, and yet produce polyketide-like products. These new enzymes, occupying an intermediate position between characterized PKS and aFAS enzymes, offer a new opportunity to better understand lipid and bioactive polyketide biosynthesis across the domains of life (2). In this work, we elucidate the core biochemical functions of this class, providing a basis to understand their key role in animal biological and chemical diversity.

Both aFAS and HRPKS consist of a set of enzyme domains that collectively perform all reactions needed to synthesize a lipid chain. The proteins act iteratively, repeatedly condensing units of malonyl-CoA to synthesize a growing lipid chain. The iteration of these precursors themselves is quite simple: they are first activated by an acyltransferase (AT) domain from malonyl-CoA, which is loaded onto the acyl carrier protein (ACP) and subsequently condensation of two malonate units takes place in the ketosynthase (KS) domain. The actions of ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) domains leads to synthesis of hydroxyl, alkene, and alkane functionality, respectively. More complex scenarios occur in some HRPKSs, in which variable

reduction or methylation by methyltransferase (MT) domain are sometimes seen (3, 4). The aFAS exhibits an additional peculiarity: an open pocket within the AT allows it to activate acetyl-CoA as a starter unit, as well as the malonyl-CoA extender unit, so that the domain is often called “MAT” (5).

Like aFAS and HRPKSs, AFPKs comprise single modules with multiple domains that act iteratively. However, the discovery of AFPKs broke several of the established rules noted above (2). The canonical AFPK, EcPKS1 from the sacoglossan mollusc *Elysia chlorotica*, builds products by the iterative addition of methylmalonyl-CoA, rather than the malonyl-CoA preferred by characterized aFAS and HRPKS. Starter-unit usage also appeared to be quite different from that found in aFAS or HRPKS. Moreover, like HRPKS but unlike aFAS, EcPKS1 exhibited variable reduction capacity, reducing some of the condensed methylmalonate units to alkene but leaving others as intact ketones.

While notable, these findings with the single EcPKS1 enzyme just scratched the surface of the structural and biochemical diversity of the AFPK products. In vitro, EcPKS1 produces **1** and **2**, which are likely precursors of the pyrone polypropionate natural products such as **3-6** found in diverse sacoglossan molluscs (Fig. 1 and *SI Appendix*, Fig. S1) (6–8). Many sacoglossans engage in kleptoplasty, stealing the plastids from algae and using them for photosynthesis (9). Because the route from **1** and **2** to the natural products involves modifications that appear to be UV- or singlet oxygen-dependent, it is possible that the compounds prevent damage that would otherwise be caused by a photosynthetic lifestyle (10–14). Products **1** and **2** do not comprise the sole precursors to sacoglossan products, since many have variable oxidation states and apparently incorporate precursors other than acetate. Moreover, other molluscs contain diverse polyketides that might also originate from AFPKs, and AFPKs are found in many other animals outside of the molluscs. Virtually nothing is known about the range of enzymes and enzymatic mechanisms needed to produce the vast array of polyketide natural products reported from animals, leading to a lack of predictability linking gene sequences to chemical structures or enabling bioengineering approaches.

Here, we sought to define the function of AFPKs by analyzing their sequences, characterizing a key AFPK, EcPKS2 from *E. chlorotica*, and then hybridizing EcPKS1 and EcPKS2 to create unnatural natural products. In our original work, we used current knowledge of sequence motifs that indicate substrate selectivity to explain the function of EcPKS1, and to predict that EcPKS2 should produce a malonate-derived polyene. Here, we show that the knowledge gained from extensive studies of other PKS classes in bacteria, animals, fungi, and plants are not predictive of AFPK function, and that new and largely undefined rules govern the broad family of AFPKs. For example, despite prediction, EcPKS2 also uses solely methylmalonate as an extender unit, indicating new substrate selectivity rules in this enzyme class. EcPKS2 accepts many different starter units, leading to classes of compounds bearing similarity to those from many different mollusc and other animal groups. Hybridization of these enzymes revealed the molecular basis of reductive regioselectivity and chain length determination.

These and other results from this study explain the origins of key chemical functional groups in AFPK products, including hundreds of natural products from molluscs; reveal an unexpected biochemical diversity of PKSs in the animal world; and demonstrate key biochemical relationships between the aFAS and HRPKS enzymes. By demonstrating that these diverse enzymes make different products, we reveal that a previously hidden wealth of biochemical diversity lies encoded in animal genomes.

Results

Sacoglossan AFPKs represent a broad family of enzymes with unknown structural diversity

AFPKs are difficult to discover because of their close sequence similarity to FASs (2). Distinguishing AFPKs from FASs necessitates thorough extensive analysis of phylogenetic trees. For example, low diversity KSs sequences of both FASs and AFPKs (fewer than 10 sequences per

type) often leads to misclassification in the phylogenetic tree. Although AFPKs are widespread in nature, they are abundant and likely produce highly structurally diverse compounds in shell-less sacoglossan molluscs (clade Plakobranchea). Therefore, to gain insight into how AFPK sequence features are reflected in chemical structures, using the method described in our previous study (2), we performed an exhaustive search of sacoglossan KS domains from 71 transcriptome SRA data sets and three genome assemblies in NCBI databases. In total, we identified 36 AFPK gene sequences from 14 species, demonstrating the diversity and ubiquity of this enzyme family in one group of molluscs (Fig. 2A). Given that EcPKS1 synthesizes polypropionate pyrones with the same scaffold, we proposed that sacoglossan pyrones are synthesized by AFPKs detected in their genomes.

We divided the sacoglossan AFPKs into subclades 1 through 9 (Fig. 2A), based upon tip distances within each subclade <1. Sacoglossans in the superfamily Plakobranchoidea are all photosynthetic by virtue of kleptoplasty and often contain long-chain polypropionate pyrones, with the chain usually cyclized to form complex rings (15, 16). Photosynthetic species expressed subclades 1 and 7-9. Subclades 8 (including EcPKS1) and 9 are especially widely shared, found in five species analyzed here. We propose that the proteins in subclades 8 and 9 likely synthesize similar polyketides, consistent with the similar polypropionates known from these species (Fig. 2B). In contrast, the sampled taxa in superfamily Limapontioidea shared a different set of AFPK lineages (subclades 2-6), although due to data constraints, insufficient proteins were discovered to link *Placida* with the primary protein cluster for related limapontioidean taxa. Most species in Limapontioidea are non-photosynthetic, and predominantly contain polypropionate pyrones of short chain length (16). The domain architecture of the enzymes in subclades 2-6 suggests that all might make the same type of polypropionates, consistent with the short-chain compounds found in non-photosynthetic sacoglossans (Fig. 2B). Interestingly, the basal plakobranchoidean *Bosellia* sp. contains proteins belonging to both subclade 5 and 8. This distribution and the phylogenetic position of *Bosellia* suggests a transition from short- to long-chain subclade enzymes at the root of the plakobranchoidean radiation, concurrent with the acquisition of kleptoplasty.

Many of the species analyzed had multiple AFPKs. We focused on *E. chlorotica* because it was the source of the characterized EcPKS1. Reanalysis of the sequencing data from GenBank (accession numbers PRJNA484060, PRJNA437117, PRJNA192596, PRJNA76673, PRJNA75367) revealed six different AFPK sequences in *E. chlorotica* (SI Appendix, Fig. S2). Their expression levels varied and were much lower than cytoplasmic FAS. The three most highly expressed genes (EcPKS2, EcPKS3 and EcPKS6) share high sequence similarity (73%-80%) and group together phylogenetically (SI Appendix, Fig. S2). Moreover, these three genes were adjacent to each other in the genome (SI Appendix, Fig. S2); the physical linkage might relate to enzyme functionality during polyketide biosynthesis or could alternatively be related to regulation of gene expression. It is possible that tandem gene duplication was involved in the early stage of AFPK diversification (17).

Previously, based upon comparison to insights from bacterial PKS sequences, we predicted that EcPKS2 in particular would synthesize a polyene derived from malonate, whereas EcPKS1 used methylmalonate (2). This was based upon an analysis of AT domains, in which the GHSXG motif in EcPKS1 was more consistent with methylmalonate, while that in EcPKS2 was more consistent with malonate, at least as found in the bacterial precedents. However, that prediction was tempered by the observation that the AT domains were quite different from those in bacteria and did not encode other diagnostic features commonly found in bacteria. Here, we therefore aimed to investigate and compare these two proteins to unveil key principles governing polyketide biosynthesis that could be used to predict function and chemical products across the diverse AFPK enzyme family.

EcPKS2 synthesizes the precursor to the sacoglossan natural product, elysione (5)

To determine whether AFPK function matched prediction, EcPKS2 (GenBank accession: QMU95646.1) was cloned and expressed in *Saccharomyces cerevisiae* BJ5464 harboring the *npaA* (phosphopantetheinyl transferase gene) and purified (SI Appendix, Fig. S3) (18). This

produced fully modified protein in which the ACP had been activated with the phosphopantetheine group. In contrast to expectation, the pure protein did not yield products unless methylmalonyl-CoA was present, indicating that malonyl-CoA was not the preferred substrate. Incubation of EcPKS2 with methylmalonyl-CoA and NADPH led to the observation of several peaks observed by UPLC-MS in negative mode (Fig. 3). One peak (compound **9**) was by far the most abundant, exhibiting a UV absorption similar to polypropionates (λ_{max} 326 nm), while two minor peaks were also present. Accurate mass measurement revealed a molecular ion $[\text{M-H}]^-$ at m/z 383.2231, indicating a molecular formula of $\text{C}_{24}\text{H}_{32}\text{O}_4$ (calcd m/z 383.2228). This formula was congruent with the incorporation of 8 methylmalonate units, but with one extra oxygen more than anticipated.

To firmly elucidate the structure of the major enzyme reaction product, we scaled up the EcPKS2 enzymatic reaction, also adding the enzyme MatB, methylmalonate, and CoA. MatB synthesizes methylmalonyl-CoA in situ to generate sufficient material for analysis in a cost-effective manner (19). Compound **9** was purified and used for spectroscopic analysis (*SI Appendix*, Table S1, Fig S9-16). From the ^1H and HSQC NMR spectra, 1 aliphatic and 7 vinylic methyl groups, 4 aromatic methines, and a methylene proton were readily distinguished. The HMBC spectrum revealed an additional 9 unsaturated carbons that were assigned to a polyene chain and a pyrone ring system. The presence of an α -pyrone was further confirmed by diagnostic chemical shifts at C-1 (δ_{C} 167.3) and C-3 (δ_{C} 166.7). An additional carbon at 214.9 ppm in the HMBC spectrum revealed an unexpected ketone oxygen. COSY and HMBC correlations firmly established the planar structure as a previously undescribed molecule, pelysione (**9**) ("precursor of elysione" (**5**)). The absolute configuration was defined using electronic circular dichroism (ECD) and NMR experiments. The ECD spectrum revealed no peaks for **9**, as expected if C-14 is racemic due to its proximity to racemizing functional groups. This is similar to precedents in other polyketides, where such protons are enzymatically synthesized in a stereospecific manner, but then racemized post synthesis (20, 21). The double bond geometries were all assigned to the *Z* configuration based upon 1D nuclear Overhauser effect (NOE) spectra; in all cases, correlations were observed to neighboring methine and methyl groups but not to cross-olefin methyl groups, indicating a *trans* relationship across double bonds (*SI Appendix*, Fig. S4).

The minor reaction products were identified as **2** and **10**, linear polyene pyrones lacking the initiating ketone group. These were present in too low an abundance for NMR characterization, with **2** and **10** together representing about 5% of the total reaction products. Nonetheless, it was possible to assign their planar structures with a high degree of confidence. High-resolution masses of **2** and **10** were $[\text{M-H}]^-$ m/z 367.2278 and 327.1975, respectively. These were consistent with molecular formulas 16 Da and 56 Da lighter than **9**, representing with high precision the lack of an oxygen in **10** in comparison to **9**, and the lack of COCH_2CH_3 for **2** in comparison to **9**. We thus hypothesized that **10** contained an additional C-C double bond in place of the C-15 ketone, while **2** lacked the terminal propionate group entirely. This hypothesis was further confirmed using $^{13}\text{C}_4$ -methylmalonyl-CoA incorporation experiments and high resolution MS^2 fragments (*SI Appendix*, Fig. S36).

When propionyl CoA was used in enzyme reactions, the chromatogram was identical to that with methylmalonate only. Thus, we hypothesized that propionyl-CoA is the starter unit. Therefore, we used experiments including $^{13}\text{C}_4$ -methylmalonyl-CoA and propionyl-CoA to confirm the incorporation. When only $\text{U-}^{13}\text{C}$ -methylmalonyl-CoA was present, ions with m/z 407.3048 (**9**), 348.2674 (**2**), and 391.3086 (**10**) were detected. When propionyl-CoA was included, these peaks were missing or very minor, and instead m/z 404.2932 (**9**), 345.2569 (**2**) and 388.2986 (**10**) (*SI Appendix*, Fig. S21) were the major ions. The difference of 3 Da resulted from a single addition of unlabeled propionyl-CoA, indicating a preference for propionate. EcPKS2 uses either methylmalonyl-CoA or propionyl-CoA as a starter unit and methylmalonyl-CoA as the extender unit to synthesize polypropionates.

Thus, in contrast to prediction, EcPKS2 produces the precursor to elysione (**5**), indicating that rules learned from decades of studying other classes of PKSs may not apply to the AFPKs. The AT domain selectivity in AFPKs is not predictable based upon current understanding of bacterial AT

domain selectivity. A particular mystery was the complex oxidation pattern of **9**, in which the first propionate unit was not reduced, the ensuing four propionates were reduced to alkenes, and the final three propionates were also not reduced. This pattern is extremely unusual in iterative PKS and was further explored in experiments described below.

EcPKS2 substrate promiscuity explains the biosynthesis of many mollusk polyketides

Most mollusk polyketides are products of methylmalonate/propionate metabolism, but there are several cases where the structural features imply that other substrates are used. For example, tridachiahydropyrone (**6**) might start from an isovalerate subunit, while placidene B (**4**) might start with acetate. Previously, it was proposed that, instead, these structural differences result from later modifications after PKS action, such as demethylation or methyl transfer (22). Other mollusk polyketides contain extender units that might originate from both methylmalonate and other extenders, such as malonate (22). For example, navenones (**7** and **8**) and related compounds appear to initiate with nicotinic acid and benzoic acid (23). To further investigate these questions, we incubated EcPKS2 with a mixture of other substrates. The resulting reaction products were analyzed by high-resolution mass spectrometry and confirmed by $^{13}\text{C}_4$ -methylmalonyl-CoA incorporation experiments. In no case could EcPKS2 synthesize products without the presence of methylmalonyl-CoA, which was required as an extender unit. However, all substrates except for malonyl-CoA were accepted as starter units. The following substrates were accepted by EcPKS2 as starter units, which in the case of malonyl derivatives were decarboxylated: acetyl-CoA, benzoyl-CoA, benzoyl-SNAC, butyryl-CoA, ethylmalonyl-CoA, isobutyryl-CoA, methylmalonyl-CoA, nicotinoyl-SNAC, and propionyl-CoA (Fig. 4, and *SI Appendix*, Fig S7-8).

Malonyl-CoA was not accepted as a starter or extender unit, and in fact inhibited product formation from methylmalonyl-CoA. Depending upon the conditions employed (especially, the relative ratios of various substrates), these starter units sometimes comprised the major enzymatic reaction products, although compound **9** was a significant component in most cases. Between 4-6 methylmalonate subunits were incorporated with this diverse set of starter units, with variable ratios of those products reflecting the relative size of the starter unit. When larger starter units were present, the C-terminal ketone products were relatively less abundant in comparison to compounds lacking that ketone. When aromatic residues pyridine or phenyl were chain initiators, the ketone was fully reduced. This likely reflects the ketoreductase (KR) selectivity, where larger initiators place the ketone further away from the chain terminus.

Overall, these results indicated that EcPKS2 is highly promiscuous in its starter unit selectivity, potentially explaining some of the natural products found in molluscs. The observed starter unit selectivity resembles that seen for aFAS, in which separate binding regions are present in the AT, one for acyl-CoAs and another for malonyl-CoA (24). It stands in contrast to the fungal HRPKSs, for which data seem to indicate a potentially more stringent starter selectivity (25, 26).

Structural characterization by MS methods

Three methods were used to characterize compounds (**2**, **9-24**) synthesized by EcPKS2. First, their accurate masses were assessed by UPLC-qTOF mass spectrometry (*SI Appendix*, Fig. S21-28). Second, the incorporation of a single starter unit was verified using the inverse labeling approach, with natural abundance starter units and [$^{13}\text{C}_4$ -methylmalonyl]-CoA (*SI Appendix*, Figures S21-28), in a manner similar to the inverse stable isotopic labeling method (27). The utility of this method was validated using NMR-characterized compound **9**. In comparison to unlabeled enzymatic reaction products, in all cases the observed masses exactly matched expectation, with all carbons comprising ^{13}C derived from methylmalonate except for those derived from starter units. Third, MS² fragments were generated by electron-activated dissociation (EAD). NMR-characterized **9** was used to establish and validate the MS² method. Pelysione (**9**) fragmented well, and the resulting ions could be matched with the structure with high accuracy. With this method in hand, the structures of the remaining compounds (**2**, **10-24**) were confirmed (*SI Appendix*, Fig. S36-37).

Thus, while multiple methods were used to support structures **2** and **10-24**, only **9** was determined with standard NMR methods, while others are not supported by a full dataset.

NADPH dependence of ketone production

The presence of a ketone at C-15 in **9** is unusual for iterative PKS biochemistry, but many sacoglossan natural products such as elysione (**5**) have a ketone in this position. Therefore, we wished to determine what biochemical factors might account for its biosynthesis. The small amount of compounds **2** and **10**, lacking a ketone at C-15, suggested that a limitation or unnatural concentration of the reductant, NADPH, might contribute to this result. Therefore, we ran enzyme assays at NADPH concentrations ranging from 0.009-50 mM (*SI Appendix*, Fig. S42). Only compound **9**, and not **2** and **10**, was detected in the presence of NADPH at low concentrations (9.38-62.5 μ M). Production of **2** and **10** remained insignificant at NADPH concentrations <1 mM, representing <1% of the reaction products (as low as 0.3%), but plateaued at about 5% (high of 7%) of the total products by 4 mM and above. At 50 mM, production of these analogs decreased, potentially due to substrate inhibition. Based upon these results, at physiological NADPH concentrations, **9** should be by far the major enzymatic product. This indicates that presence of a compound **9** with a ketone at C-15 is the true product formed by the AFPK in nature, and does not result from modification after biosynthesis or as an artifact of insufficient NADPH.

Substrate selectivity comparison with EcPKS1

Previous work with EcPKS1 showed that it could not accept acetyl- or malonyl-CoA, but instead used methylmalonyl-CoA exclusively. By contrast, EcPKS2 could accept every substrate attempted except for malonyl-CoA. In initial studies, we repeated the experiments and found that, similarly, EcPKS1 only accepted methylmalonyl-CoA. In comparison to EcPKS2, EcPKS1 provided a relatively much lower product yield, leading us to suspect that the holoprotein purified from yeast might be inhibited in comparison to EcPKS2. Thus, we expressed the EcPKS1 apo form in a yeast strain lacking NpgA, and then pantetheinylated the ACP with the purified apoprotein in vitro using Sfp (28). To our surprise, we found that the in vitro synthesized apoprotein was broadly accepting of starter units, including methylmalonyl-, propionyl-, isovaleryl-, and benzoyl-CoA. We could even detect a slight trace of products derived from acetyl-/malonyl-CoA as the starter unit. This led us to hypothesize that EcPKS1 also does not discriminate between acetate versus propionate starter units, like EcPKS2. Instead, we proposed that a discrete intermediate incorporating an acetate starter unit inhibited product formation. Since yeast cells are dominated by acetyl- and malonyl-CoA, this would potentially inhibit an enzyme that was active in vivo.

To test this hypothesis, we incubated apoEcPKS1, lacking the phosphopantetheinylation required to activate the ACP, with CoA substrates. Previous studies show that PKSs can directly utilize such active substrates even in the absence of a functional ACP, and that such experiments enable detection of early steps in the reaction cycle (29). Here, apoEcPKS1 was treated with either acetyl- or malonyl-CoA, and methylmalonyl-CoA in the presence of NADPH. Abundant tiglyl-CoA (**34**) was obtained (*SI Appendix*, Fig. S38), representing the activity of the KS in condensing one unit of acetate and one unit of methylmalonate, followed by ketoreduction and dehydration. However, tiglyl-CoA was not further chain extended. Similarly, the same conditions absent NADPH led to the production of the homologous compound (**35**), with ketone still intact (*SI Appendix*, Fig. S41). The structures of **34** and **35** were assigned by high-resolution MS data, coupled with generation of a standard for **34**.

By contrast, when apoEcPKS1 was incubated with methylmalonyl-CoA and NADPH, no 2-methyl-2-pentenal-CoA (**36**) accumulated, and instead it appeared that substrates were further elongated by the enzyme (*SI Appendix*, Fig. S40). Thus, we proposed that tiglyl-ACP inhibits the function of EcPKS1 because it is not a good substrate for further chain extension in the KS. In summary, the

substrates used by EcPKS1 and EcPKS2 are similar, except that EcPKS1 did not accept the intermediate tiglyl-CoA for further chain elongation.

Hybrid enzymes explain the product suite synthesized by EcPKS1 and EcPKS2

The experiments described above left several immediate questions: what controls the regioselectivity of ketone reduction in EcPKS2? Is the KS domain really responsible for tiglic acid inhibition of EcPKS1, and what controls substrate selectivity? What controls the chain length differences observed in EcPKS1 versus EcPKS2 products, and potentially the larger family of sacoglossan polyketides? We envisioned that answering these questions would enable better sequence-function prediction and engineering of AFPKs, while also potentially shedding light on aFAS and iterative PKS function.

We synthesized hybrid enzymes in which either the KS or the KS-AT didomain from one protein was fused with the DH-MT⁰-ER⁰-KR-ACP domains from another. The KS-only fusions proved difficult to work with, being very poorly soluble. By contrast, soluble KS-AT fusions could easily be expressed. The fusion between EcPKS1 KS-AT and EcPKS2 DH-MT⁰-ER⁰-KR-ACP is hereafter referred to as “EcPKSf1-2”, while its counterpart with the EcPKS2 KS-AT and EcPKS1 DH-MT⁰-ER⁰-KR-ACP is “EcPKSf2-1”. Since these proteins are only ~52% sequence identical, optimization of conditions was required for the expression and purification of mutants. In comparison to wild type, in the mutants a greater percentage was found in the aggregate fraction rather than in the functional dimeric fraction, as indicated by a smaller proportion of soluble dimeric protein observed in size-exclusion chromatograms (*SI Appendix*, Fig. S3). The purified, dimeric protein was used in all experiments.

In answering the question about regioselectivity of ketone reduction, we envisioned two possibilities: control of this modification may be the result of either the KS or KR domains (or potentially, a combination of both). For example, if KS were responsible, the EcPKS1 KS domain would potentially not extend a ketone-containing starter unit. If the KR were responsible, then the KR from EcPKS2 would govern ketone formation regardless of whether the KS from EcPKS1 or EcPKS2 was present. When the fusion proteins were treated with methylmalonyl-CoA, EcPKSf1-2 gave primarily compound **25**, in which a ketone was present in the first oxidation position, whereas EcPKSf2-1 gave primarily **2** and **10**, in which the ketone was absent. This result unambiguously demonstrated that the reductive domains, and specifically the KR_s in the AFPKs, controls the complex reduction regioselectivity. In addition, EcPKSf1-2 gave shunt products **30** and **31**. We believe that this may represent imperfect complementation between the ACPs and KS domains of EcPKS1 and EcPKS2, since the intermediates resemble what was observed in apo-protein experiments.

These results also strongly supported the role of KS in chain length determination, independent of the KR. Compounds **2** and **10** (with a chain length closer to the native products of EcPKS2) produced by EcPKSf2-1 reflect the chain length preferences of EcPKS2, even though the ketone is absent; similarly, **25** (with a chain length equal to the native products of EcPKS1) produced by EcPKSf1-2 is the preferred chain length of EcPKS1, but with a carbonyl present in the first propionate group. In all cases, results were as expected if KS controls the chain length.

To determine whether tiglyl-CoA (**34**) is really likely to inhibit the KS of EcPKS1, we repeated the same experiments described above for apoEcPKS enzymes. As found with apoEcPKS1, abundant tiglyl-CoA was detected based upon high resolution Q-ToF data in apoEcPKSf1-2, but not in EcPKSf2-1 (*SI Appendix*, Fig. S38). The results were identical, demonstrating one turnover through the KS and one through KR/DH, strongly supporting our identification of tiglate as an inhibitor of chain elongation by the KS (*SI Appendix*, Fig. S39). Thus, overall, the KS role in substrate selectivity and chain length was reinforced by hybridization experiments.

Discussion

The AFPKs represent a new class of enzymes encoded in the genomes of phylogenetically diverse animal lineages, and which bridge the evolutionary gap between aFAS and HRPKSs (2). While it

is now well established that animals have their own groups of PKSs as well, the animal PKS enzymes function much differently from the aFASs, and they are more distantly related to aFASs than are the fungal HRPKSs (30). Extensive studies of biosynthetic enzymes such as the PKSs found in bacteria and fungi have yielded an integrative knowledge base connecting molecular sequence to protein function to chemical structure. These principles have led to key enabling technologies that now make it much easier to go from DNA sequence to chemical structure, and to perform engineering of unnatural natural products and drug discovery studies (31, 32).

No such knowledge base exists for animal PKSs, let alone for the newly defined AFPKs (2, 33, 34). It is currently impossible to predict the products produced by the countless PKS and AFPK enzymes found in animals, and engineering of these pathways has yet to be performed or even proposed in the literature. Here, we aimed to remedy some of these key limitations by characterizing EcPKS2 and comparing its function with the related enzyme, EcPKS1. We envisioned that these studies would provide a foundation for understanding how molecular features of the biosynthetic machinery explain the many polyketide-like chemical structures found in diverse animals. Simultaneously, we would learn which protein domains control key factors such as regioselectivity and chain length, enabling structure prediction and engineering approaches.

Here, we show that AFPKs combine key structural diversity-generating principles found in aFAS and in HRPKSs. For example, like aFASs, EcPKS1 and EcPKS2 accept a broad range of starter units (Figure 6). This may be due to a similarity in the AT between the aFASs and AFPKs (35, 36). Alternatively, in bacterial modular PKSs, the KS domain is implicated as the gatekeeper governing starter selectivity (37, 38). Like HRPKSs, instead of being restricted to repetitive, saturated alkanes, the AFPKs produce a complex oxidation pattern, where the oxidation state varies across the length of the polyketide (39). Together, these two properties would enable the production of most of the diverse polyketides found in sacoglossans and other molluscs. AFPKs also exhibit their own unique properties. Unlike HRPKSs and aFASs that are highly restricted to using malonate, EcPKS1 and EcPKS2 utilize methylmalonate. In addition, the complex pattern of ketone reduction observed in the EcPKS2 products is quite different than that observed in HRPKSs so far studied. Such differences observed in just two of the vast number of AFPKs hints at much biochemical and structural diversity awaiting discovery.

Our sequence analysis revealed a phylogenetic tree of AFPK genes reflecting the different product suites found in photosynthetic versus nonphotosynthetic sacoglossans. Since the long-chain polyketides are associated with photoprotection and photosynthesis, the products of AFPKs potentially serve a different biological role than those of other animal PKSs (14, 40–42). Our biochemical results may also explain the broad suite of molluscan natural products: the different starter units and oxidation state patterns found in many molluscan polyketides are recapitulated in the products created in this study.

The biosynthetic steps governed by EcPKS1 and EcPKS2 can be proposed as follows. In both enzymes, the starter unit is first generated either by loading the ACP with methylmalonate followed by decarboxylation, or by loading an acyl-thioester onto the KS. Subsequent condensation with ACP-bound methylmalonate yields the first diketide intermediate. In EcPKS1, this intermediate is reduced to the α , β -unsaturated thioester, while it is not modified by EcPKS2; the redox difference is solely governed by the KR domain. Next, the diketide is extended by either 1 (EcPKS1) or 3 (EcPKS2) methylmalonate units, in which the ketones are reduced and dehydrated to the double bonds. Finally, an additional 3 methylmalonate units are added without reduction and spontaneously released from the ACP via pyrone formation (*SI Appendix*, Fig. S43). After the action of the PKSs, other as-yet unidentified enzymes are likely involved in further processing toward the complex natural products. Since these enzymes do not appear to be encoded in gene clusters along with the PKSs, they are not described in the current study. Interestingly, there is the potential that at least some of the EcPKS enzymes are encoded in a biosynthetic gene cluster, since they are found in adjacent positions in the *E. chlorotica* genome.

While this current study was extensive, it leaves many remaining mysteries about EcPKS1 and EcPKS2. The enzymes decarboxylate methylmalonyl-CoA to generate starter units; how this is accomplished, in comparison to aFASs which do not perform this reaction, has yet to be characterized. We do not yet know conclusively whether the promiscuous starter unit selectivity of these enzymes is governed by KS, as found in some bacterial PKSs, or by AT, as found in aFAS. The strict selectivity for methylmalonyl-CoA extender units might be determined by the AT or KS domains, or a combination of both. The release mechanism from the ACP is still not understood, and potentially may include the KS domain. Finally, despite predictive efforts to categorize the sacoglossan AFPKs, we still do not have predictive power to anticipate function based upon sequence. Given the many AFPKs found in nature, it is important to address these limitations to better understand the breadth of lipid diversity in the animal kingdom.

Although little is known about pyrone biosynthesis in animals, pyrones themselves are common products or byproducts of PKS enzymology. In an interesting case of convergent evolution, the EcPKS2 products bear similarity to bacterial polypropionate pyrones such as aureothin (43). Unlike EcPKS2 that uses a single module iteratively, enzymes that make aureothin and relatives are modular, comprising a different module for each unit loaded (44–46). An exception is the first module, which loads the starter unit and extends by two methylmalonyl units, yet still operates by some different rules in comparison to EcPKS2. The loading module in aureothin biosynthesis was flexible, enabling the creation of analogs (47). It is difficult to compare biochemical studies directly, but from our data it appears that EcPKS2 may be more efficient in processing a diverse array of substrates. It would be interesting to hybridize these two biochemical systems to see whether increased biochemical diversity is feasible in the broad chemical class of pyrones. The biosynthetic plasticity of EcPKS2 should lend itself to a range of applications including bioengineering and synthetic biology efforts.

This study represents one of only a few biochemical studies on animal PKS or AFPK enzymes (2, 33, 34). In each case so far reported, significantly different chemical products and biochemical features have been seen, reflecting an immense potential of the unstudied animal AFPKs and PKSs to produce new compounds with unanticipated biochemical reactions. Overall, these results map the biosynthesis of animal polyketides, connecting mollusc genetics and enzymology with the resulting structural diversity of their natural products. Because many more AFPKs are found in molluscs and await characterization, we expect to see further structural diversity unveiled from this unusual group of PKS-like enzymes.

Materials and Methods

Detailed Materials and Methods are described in the *SI Appendix*.

KS-containing protein identification. The raw transcriptome sequencing data for sacoglossan molluscs was downloaded from SRA database and assembled using rnaSPADES. The genes were called out using prodigal using “-meta” mode. Previously identified animal FAS, AFPK and animal PKS sequences were used as query to do Blastp search against the transcriptome database. The target sequences that formed significant alignment (qcov > 80) with the queries were considered as PKS hits. To remove any contamination from the SRA transcriptome assemblies, the corresponding contigs that contain the KS hits were analyzed using the taxonomy assignment pipeline in the Autometa package (make_taxonomy_table.py -a ks_hit_contigs.fa -l 700) (43). The output ‘.lca’ file gives taxonomy ID for the lowest common ancestor of each contig. Based on the taxonomy ID, contigs for bacteria, fungi, plants, and algae were removed. The KS domains of the remaining contigs were predicted by antiSMASH and InterPro. Protein sequences for *Elysia chlorotica* were also downloaded from NCBI (RQTK01) and analyzed using the same method above to obtain the KS-containing protein sequences.

Phylogenetic analyses. Orthologous genes were aligned using t-Coffee (-mode mcoffee -output = msf, fasta_aln) (48). To remove poorly aligned regions, the resulting alignment was subsequently trimmed with Clipkit with model parameter ‘-m kpi-gappy’ (49). The trimmed alignment was then manually inspected to remove any remaining poorly aligned region. The maximum likelihood tree

was constructed using iqtree (/iqtree -nt AUTO -st AA -alrt 1000 -bb 1000). (50) The ML tree was visualized using the ggtree library (51)

Overexpression and purification of PKSs. Expression vector *pxw55-EcPKS1*, *pxw55-EcPKS2*, *pxw55-EcPKSf1-2* and *pxw55-EcPKSf2-1* were transformed into *Saccharomyces cerevisiae* BJ5464-*NpgA* (MAT α *ura3-52 trp1 leu2- Δ 1 his3 Δ 200 pep::HIS3 prb1d1.6R can1 GAL*, with phosphopantetheinyl transferase *npgA*) or *Saccharomyces cerevisiae* BJ5464 (18, 52). Single colonies were propagated in liquid broth (1 mL) and then transferred to yeast peptone dextrose broth (1 L, 10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose) in 2.8 L Fernbach flasks. Expression was performed at 28 °C at 180 rpm for 3 d. Proteins were purified by Ni-NTA and further chromatography methods including FPLC.

In vitro AFPK assays. Purified enzyme (5 μ M) was incubated with MatB (19) (5 μ M), Sfp (28) (5 μ M), MgCl₂ (7 mM), ATP (7 mM), methylmalonate (5 mM), coenzyme A (2 mM), and NADPH (2 mM), in aqueous sodium phosphate (25 μ L; 100 mM, pH 7.5) at 22 °C for 18 h, keeping reactions in the dark.

Purification of EcPKS2 reaction products. A 40 mL-scale reaction mixture contained EcPKS2 (200 mg), sodium phosphate (100 mM, pH 7.5), MatB (70 mg), Sfp (20 mg), MgCl₂ • 6 H₂O (50 mg), ATP (150 mg), methylmalonic acid (20 mg), coenzyme A (50 mg), and NADPH (50 mg). After 2 h at 22 °C, an additional aliquot (10 mL of 100 mM sodium phosphate pH 7.5 containing ATP (70 mg), methylmalonic acid (20 mg), coenzyme A (10 mg), and NADPH (50 mg)) was added at a rate of 0.1 mL/min. The resulting mixture was stirred in the dark for 20 h at 22 °C. The reaction was quenched with ethyl acetate (50 mL), and the mixture was centrifuged for 30 min at 4500 $\times$ g to separate precipitated protein and liquid. The ethyl acetate layer was removed, and the water layer was further extracted twice more with ethyl acetate (50 mL). The combined ethyl acetate fraction was dried with sodium sulfate, concentrated under reduced pressure, and purified using HPLC (Thermo Scientific, UltiMate 3000) with Luna 5 μ m Phenyl-Hexyl 100 Å LC column (250 $\times$ 10 mm) with a linear gradient of 30-80% mobile phase B over 30 min (solvent A, H₂O; solvent B, MeCN) at a flow of 3.5 mL/min; *t_R*=25.6 min; yield of isolated compound **9**: 0.5 mg.

In vitro reaction to generate holoenzymes. Purified apo-AFPK (5 μ M), Sfp (20 μ M), MgCl₂ (10 mM), coenzyme A (5 mM) were combined in buffer (100 mM sodium phosphate pH 7.5) at 22 °C for 2 h and used in described enzymatic reactions.

¹³C₄-methylmalonyl-CoA labeling experiments. AFPK-enzymes (5 μ M), ¹³C₃-methylmalonic acid (3 mM), MatB (5 μ M), starter units (4 mM; except benzoyl-SNAC and nicotinoyl-SNAC at 16 mM), MgCl₂ (7 mM), ATP (7 mM), and NADPH (2 mM) were added to sodium phosphate buffer (25 μ L; 100 mM, pH 7.5). Starter units included ethylmalonyl-CoA, butyryl-CoA, isobutyryl-CoA, isovaleryl-CoA, acetyl-CoA, benzoyl-CoA, benzoyl-SNAC, or nicotinoyl-SNAC. The mixtures were kept in the dark at 22 °C for 18 h.

Chemical analysis. LCMS used an Acquity UPLC protein BEH C4 1.7 μ m column (2.1 $\times$ 100 mm) with a 6 min linear gradient of 5-100% mobile phase B (solvent A, H₂O 0.05 % formic acid; solvent B, MeCN) at 0.4 mL/min. Sample (5 μ L) was injected using an autosampler. MS method: negative ion and sensitivity mode; range: 100-1000 Da; scan time: 0.5 s; data format: centroid. MS/MS method: negative ion and sensitivity mode; range: 100-1000 Da, scan time: 0.1 s; data format: continuum; fixed collision energy value at 30 V.

Data Availability. All study data are included in the article and/or *SI Appendix*.

Acknowledgments

This work was funded by NSF CHE 2203613. Phylogenetic analysis was funded by NSF IOS 2127111. We thank P. D. Scesa and J. Skalicky for assistance with NMR data acquisition and J.A. Maschek for assistance with partial LC-MS data acquisition. *S. cerevisiae* strain BJ5464 and plasmid *pxw55* were gifts of Nancy Da Silva (UC Irvine) and Yi Tang (UCLA), respectively.

References

1. B. Chakravarty, Z. Gu, S. S. Chirala, S. J. Wakil, F. A. Quiocho, Human fatty acid synthase: Structure and substrate selectivity of the thioesterase domain. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15567–15572 (2004).
2. J. P. Torres, Z. Lin, J. M. Winter, P. J. Krug, E. W. Schmidt, Animal biosynthesis of complex polyketides in a photosynthetic partnership. *Nat. Commun.* **11**, 2882 (2020).
3. C. T. Walsh, Y. Tang, *Natural Product Biosynthesis* (The Royal Society of Chemistry, 2017).
4. C. Hertweck, The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed.* **48**, 4688–4716 (2009).
5. A. Rittner, K. S. Paithankar, K. V. Huu, M. Grininger, Characterization of the Polyspecific Transferase of Murine Type i Fatty Acid Synthase (FAS) and Implications for Polyketide Synthase (PKS) Engineering. *ACS Chem. Biol.* **13**, 723–732 (2018).
6. G. Cimino, M. L. Ciavatta, A. Fontana, M. Gavagnin, *Bioactive Compounds from Natural Sources* (Tringali, C. Ed, 2001).
7. Z. H. Chen, Y. W. Guo, X. W. Li, Recent advances on marine mollusk-derived natural products: chemistry, chemical ecology and therapeutical potential. *Nat. Prod. Rep.* **40**, 509–556 (2023).
8. A. Cutignano, G. Villani, A. Fontana, One Metabolite, Two Pathways: Convergence of Polypropionate Biosynthesis in Fungi and Marine Molluscs. *Org. Lett.* **14**, 992–995 (2012).
9. K. Händeler, Y. P. Grzybowski, P. J. Krug, H. Wägele, Functional chloroplasts in metazoan cells - A unique evolutionary strategy in animal life. *Front. Zool.* **6**, 1–18 (2009).
10. C. Ireland, P. J. Scheuer, Photosynthetic Marine Mollusks: In vivo ¹⁴C Incorporation into Metabolites of the Sacoglossan Placobranchus ocellatus. *Science (80)*. **205**, 922–923 (1979).
11. C. Ireland, J. Faulkner, The metabolites of the marine molluscs Tridachiella diomedea and Tridachia crispata. *Tetrahedron* **37**, 233–240 (1981).
12. A. K. Miller, D. Trauner, Mining the Tetraene Manifold: Total Synthesis of Complex Pyrones from Placobranchus ocellatus. *Angew. Chem. Int. Ed.* **44**, 4602–4606 (2005).
13. S. Brückner, J. E. Baldwin, J. Moses, R. M. Adlington, A. R. Cowley, Mechanistic evidence supporting the biosynthesis of photodeoxytridachione. *Tetrahedron Lett.* **44**, 7471–7473 (2003).
14. D. R. Zuidema, A. K. Miller, D. Trauner, P. B. Jones, Photosensitized conversion of 9,10-deoxytridachione to photodeoxytridachione. *Org. Lett.* **7**, 4959–4962 (2005).
15. J. Mistry, R. D. Finn, S. R. Eddy, A. Bateman, M. Punta, Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**, e121 (2013).
16. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
17. M. Morley, *et al.*, Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743–747 (2004).
18. R. A. Cacho, Y. Tang, Reconstitution of Fungal Nonribosomal Peptide Synthetases in Yeast and In Vitro. *Methods Mol. Biol.* **1401**, 103–119 (2016).
19. A. J. Hughes, A. Keatinge-Clay, Enzymatic extender unit generation for in vitro polyketide synthase reactions: Structural and functional showcasing of Streptomyces coelicolor MatB. *Chem. Biol.* **18**, 165–176 (2011).
20. B. Sedgwick, J. W. Cornforth, The Biosynthesis of Long-Chain Fatty Acids. Stereochemical Differentiation in the Enzymic Incorporation of Chiral Acetates. *Eur. J. Biochem.* **75**, 465–479 (1977).
21. K. J. Weissman, *et al.*, The molecular basis of Celmer's rules: The stereochemistry of the condensation step in chain extension on the erythromycin polyketide synthase. *Biochemistry* **36**, 13849–13855 (1997).
22. V. Di Marzo, *et al.*, Cyercenes, novel pyrones from the ascoglossan mollusc Cyerce cristallina. Tissue distribution, biosynthesis and possible involvement in defense and regenerative processes. *Experientia* **47**, 1221–1227 (1991).

23. Howard L. Sleeper and William Fenical, Navenones A-C: trail-breaking alarm pheromones from the marine opisthobranch *Navanax inermis*. *J. Am. Chem. Soc.* **99**, 2367–2368 (1977).
24. A. Rittner, K. S. Paithankar, A. Himmler, M. Grninger, Type I fatty acid synthase trapped in the octanoyl-bound state. *Protein Sci.* **29**, 589–605 (2020).
25. D. A. Herbst, C. A. Townsend, T. Maier, The architectures of iterative type I PKS and FAS. *Nat. Prod. Rep.* **35**, 1046–1069 (2018).
26. C. Chiang, M. Ohashi, Y. Tang, Deciphering chemical logic of fungal natural product biosynthesis through heterologous expression and genome mining. *Nat. Prod. Rep.* **40**, 89–127 (2023).
27. D. A. Cummings, A. I. Snelling, A. W. Puri, Methylotroph Quorum Sensing Signal Identification by Inverse Stable Isotopic Labeling. *ACS Chem. Biol.* **16**, 1332–1338 (2021).
28. Z. Zhou, *et al.*, Genetically Encoded Short Peptide Tags for Orthogonal Protein Labeling by Sfp and AcpS Phosphopantetheinyl Transferases. *Am. J. Chem. Biol.* **2**, 337–346 (2007).
29. O. Ad, B. W. Thuronyi, M. C. Y. Chang, Elucidating the mechanism of fluorinated extender unit loading for improved production of fluorine-containing polyketides. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E660–E668 (2017).
30. T. A. Castoe, T. Stephens, B. P. Noonan, C. Calestani, A novel group of type I polyketide synthases (PKS) in animals and the complex phylogenomics of PKSs. *Gene* **392**, 47–58 (2007).
31. M. Klaus, M. Grninger, Engineering strategies for rational polyketide synthase design. *Nat. Prod. Rep.* **35**, 1070–1081 (2018).
32. A. Rittner, *et al.*, Chemoenzymatic synthesis of fluorinated polyketides. *Nat. Chem.* **14**, 1000–1006 (2022).
33. F. Li, *et al.*, Sea Urchin Polyketide Synthase SpPks1 Produces the Naphthalene Precursor to Echinoderm Pigments. *J. Am. Chem. Soc.* **144**, 9363–9371 (2022).
34. T. F. Cooke, *et al.*, Genetic Mapping and Biochemical Basis of Yellow Feather Pigmentation in Budgerigars. *Cell* **171**, 427–439 (2017).
35. S. Smith, A. Stern, The Effect of Aromatic CoA Esters on Fatty Acid Synthetase : Biosynthesis of w-Phenyl Fatty Acids. *Arch. Biochem. Biophys.* **222**, 259–265 (1983).
36. A. Witkowski, A. K. Joshi, S. Smith, Characterization of the Interthiol Acyltransferase Reaction Catalyzed by the β -Ketoacyl Synthase Domain of the Animal Fatty Acid Synthase. *Biochemistry* **36**, 16338–16344 (1997).
37. M. Hirsch, B. J. Fitzgerald, A. T. Keatinge-Clay, How cis-Acyltransferase Assembly-Line Ketosynthases Gatekeep for Processed Polyketide Intermediates. *ACS Chem. Biol.* **16**, 2515–2526 (2021).
38. L. Zhang, *et al.*, Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. *Angew. Chem. Int. Ed.* **56**, 1740–1745 (2017).
39. Y. T. Eun Bin Go, *Fungal Highly Reducing Polyketide Synthases and Associated Natural Products, Comprehensive Natural Products III: Chemistry and Biology*, (2020).
40. D. R. Zuidema, P. B. Jones, Triplet photosensitization in cyercene A and related pyrones. *J. Photochem. Photobiol. B Biol.* **83**, 137–145 (2006).
41. D. R. Zuidema, P. B. Jones, Photochemical relationships in Sacoglossan polypropionates. *J. Nat. Prod.* **68**, 481–486 (2005).
42. J. de Vries, *et al.*, Comparison of sister species identifies factors underpinning plastid compatibility in green sea slugs. *Proceedings. Biol. Sci.* **282** (2015).
43. N. N. Gerber, Phenazines, phenoxazinones, and dioxopiperazines from *Streptomyces thioluteus*. *J. Org. Chem.* **32**, 4055–4057 (1967).
44. J. He, C. Hertweck, Functional Analysis of the Aureothin Iterative Type I Polyketide Synthase. *ChemBioChem* **6**, 908–912 (2005).
45. N. Traitcheva, H. Jenke-kodama, J. He, E. Dittmann, C. Hertweck, Non-Colinear Polyketide Biosynthesis in the Aureothin and Neoaureothin Pathways : An Evolutionary Perspective. *ChemBioChem* **8**, 1841–1849 (2007).

46. B. Busch, C. Hertweck, Phytochemistry Evolution of metabolic diversity in polyketide-derived pyrones : Using the non-colinear aureothin assembly line as a model system. *Phytochemistry* **70**, 1833–1840 (2009).
47. M. Werneburg, *et al.*, Exploiting Enzymatic Promiscuity to Engineer a Focused Library of Highly Selective Antifungal and Antiproliferative Aureothin Analogues. *J. Am. Chem. Soc.* 10407–10413 (2010).
48. C. Notredame, D. G. Higgins, J. Heringa, T-coffee: a novel method for fast and accurate multiple sequence alignment¹¹Edited by J. Thornton. *J. Mol. Biol.* **302**, 205–217 (2000).
49. J. L. Steenwyk, T. J. Buida III, Y. Li, X.-X. Shen, A. Rokas, ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLOS Biol.* **18**, 1–17 (2020).
50. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
51. L.-G. Wang, *et al.*, Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. *Mol. Biol. Evol.* **37**, 599–603 (2020).
52. K. K. M. Lee, N. A. D. Silva, J. T. Kealey, Determination of the extent of phosphopantetheinylation of polyketide synthases expressed in *Escherichia coli* and *Saccharomyces cerevisiae*. *Anal. Biochem.* **394**, 75–80 (2009).

Figure and Table Legends

Figure 1. AFPKs and their products. (A) The domain architecture of EcPKS1 and EcPKS2, products of EcPKS1 (**1** and **2**), and representative polyketides isolated from mollusks (**3-8**). (B) Phylogenetic analysis reveals that AFPKs are very similar to animal FAS and only distantly related to PKSs. ⁰ indicates an inactive domain.

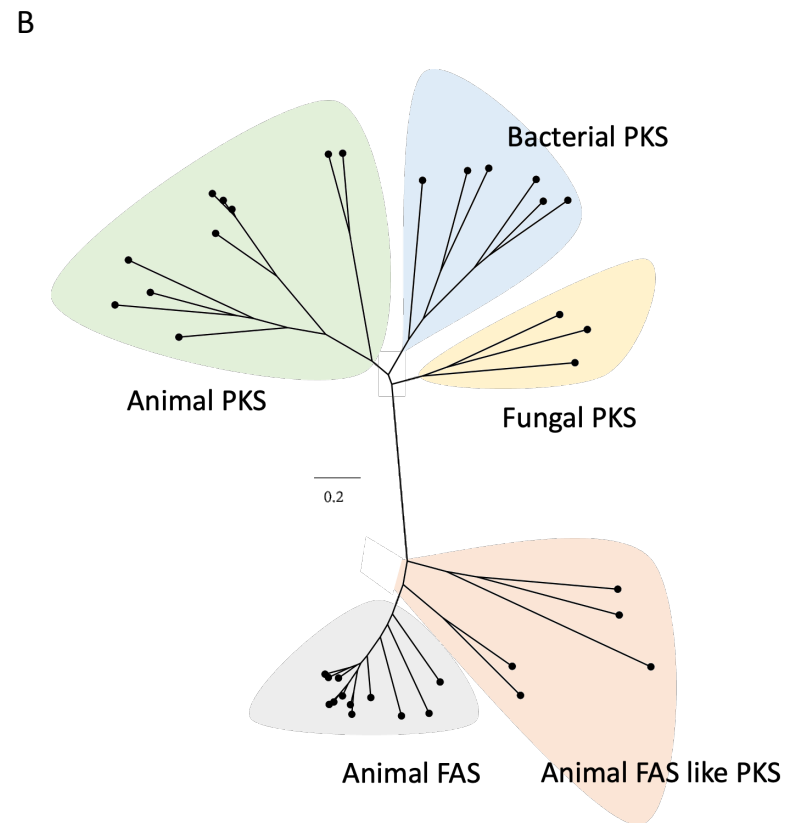
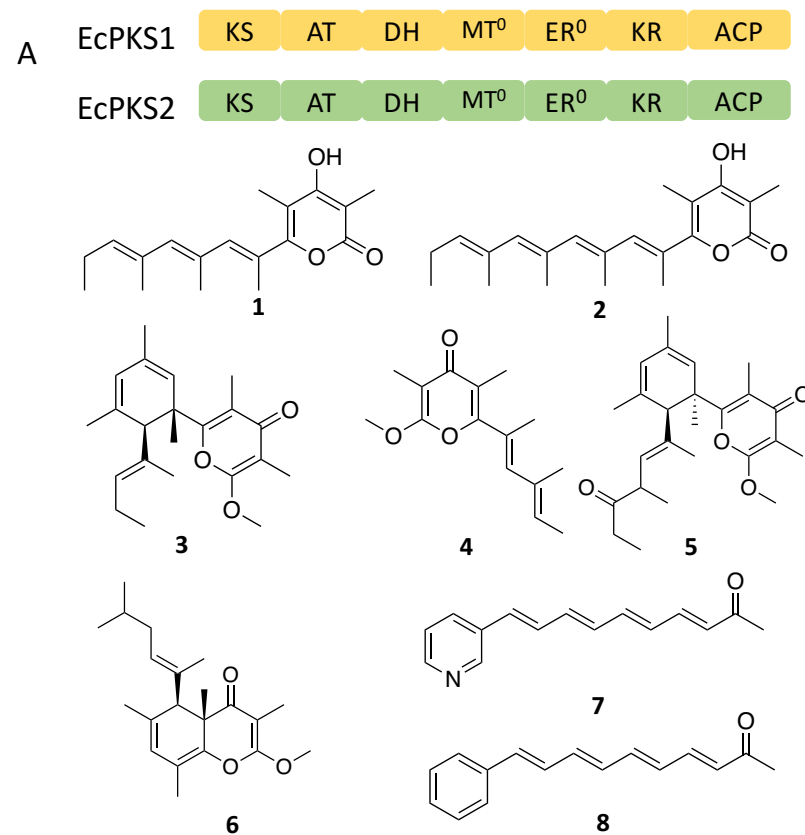
Figure 2. AFPK distribution in sacoglossan molluscs. (A) Phylogenetic tree of sacoglossan AFPKs, demonstrating that they form nine separate subclades (see Fig. S4D for a larger version of this tree). (B) Comparison of AFPK KSs with polyketide chain length in the sacoglossans. Molluscs at right often contain longer-chain products, reflected in the AFPK phylogeny, while those at left often contain shorter-chain products.

Figure 3. EcPKS2 synthesizes polypropionate precursors. UPLC-MS chromatograms are shown, with the x-axis as time in min and the y-axis being ion counts, analyzed using method *a* (*SI Appendix*). (A) EcPKS2 synthesizes the elysione (**5**) precursor (**9**) in the presence of 2 mM methylmalonyl-CoA and 2 mM NADPH. The inset shows the same result, but in the presence of additional propionyl-CoA (2 mM). (B) EcPKS2 synthesizes tridachiahidropyrone precursors in the presence of 2 mM isovaleryl-CoA, 2 mM methylmalonyl-CoA, and 2 mM NADPH. While **9** was fully characterized by NMR, the other compounds in this figure were present in low abundance or were less stable, and thus were assigned using isotope labeling experiments and a novel application of electron activated dissociation (EAD) MS, which provides a large number of fragment ions for statistical analysis (see *SI Appendix*). The small numbers present on each compound indicate the chain length from the initiating carboxyl group.

Figure 4. EcPKS2 has promiscuous selectivity for the starter unit. EcPKS2 reacts with 2 mM NADPH, 2 mM methylmalonyl-CoA, and: (A) 2 mM ethylmalonyl-CoA, and the inset shows the same result when butyryl-CoA^a (2 mM) replace ethylmalonyl-CoA; (B) 2 mM acetyl-CoA^a; (C) 2 mM isobutyryl-CoA^a; (D) 2 mM benzoyl-CoA^a (inset shows peaks that are closely overlapping in the main figure); (E) 16 mM benzoyl SNAC^b (inset shows peaks that are closely overlapping in the main figure); (F) 16 mM nicotinoyl SNAC^b. UPLC-MS chromatograms are shown, with the x-axis as time in min and the y-axis as ion counts. ^a and ^b indicate different chemical analysis methods described in *SI Appendix*. Small numbers on structures are the carbon position indicating chain length.

Figure 5. HPLC-MS chromatograms showing the major products when EcPKSf2-1 and EcPKSf1-2 are treated with methylmalonyl-CoA and NADPH.

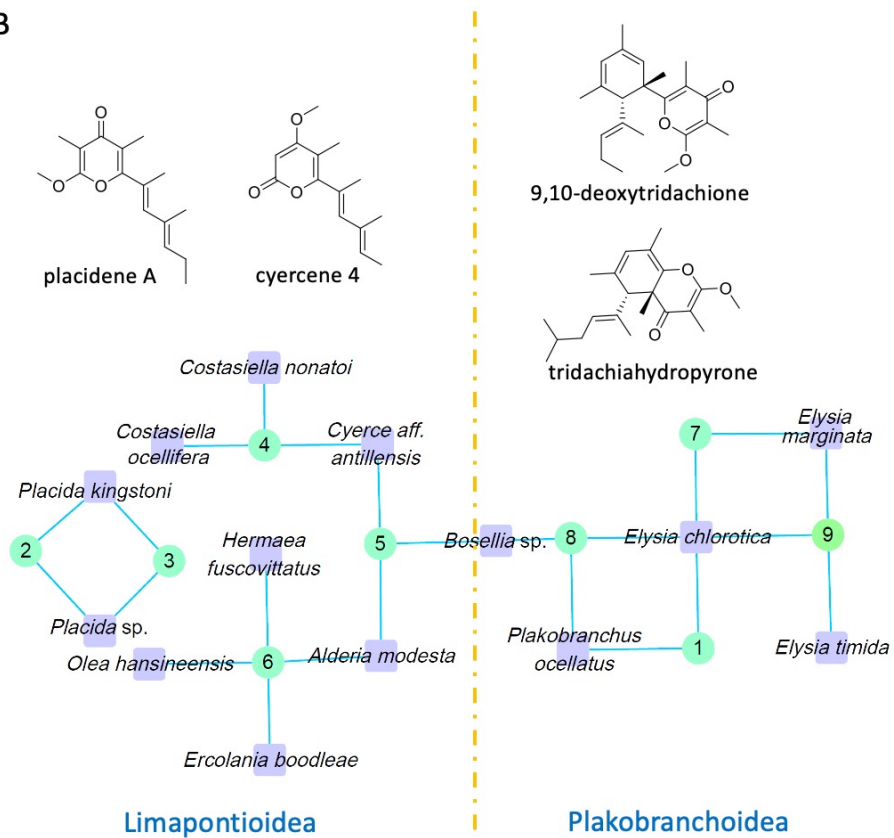
Figure 6. Representative products from enzyme assays.

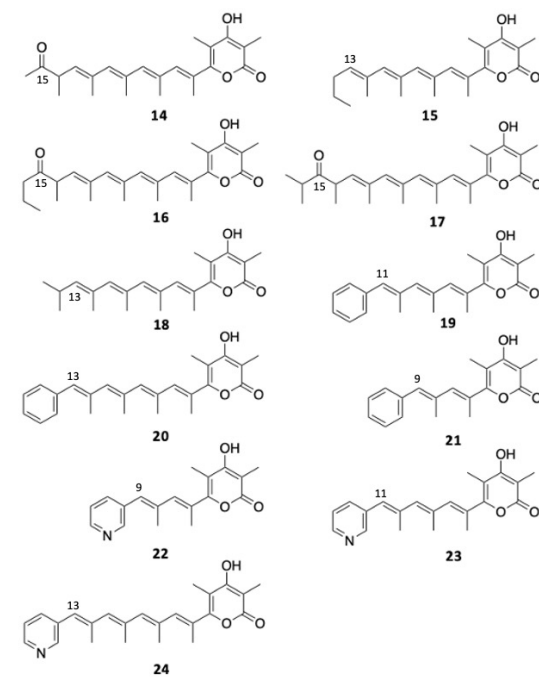
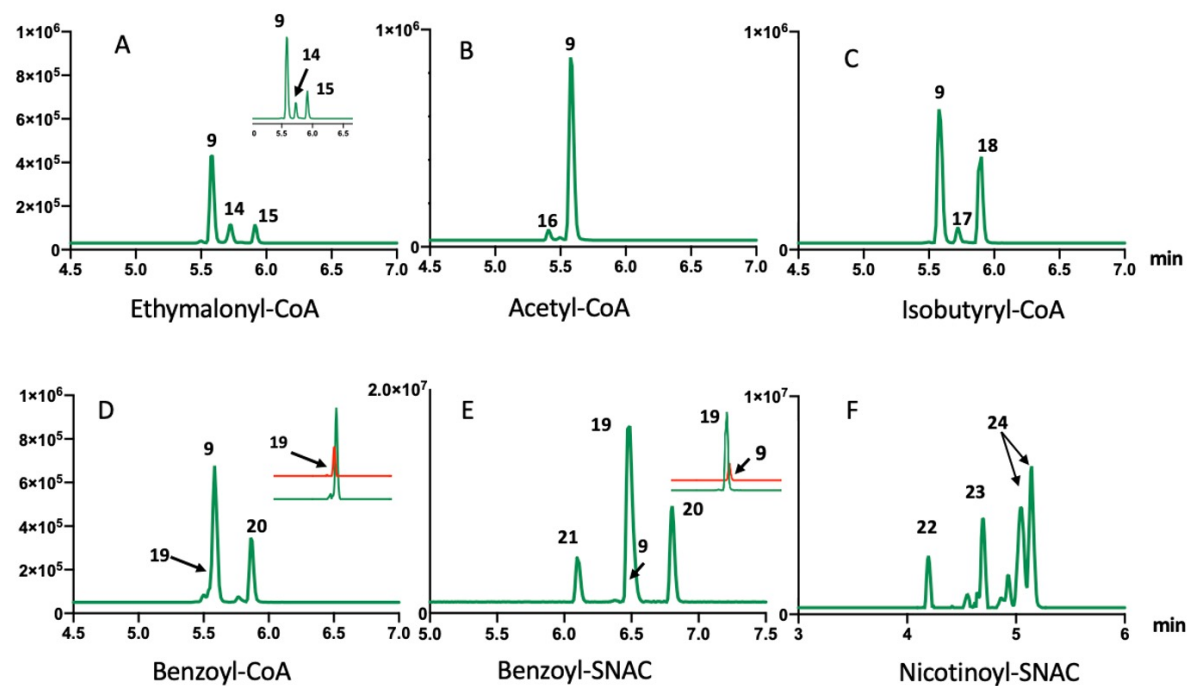


A



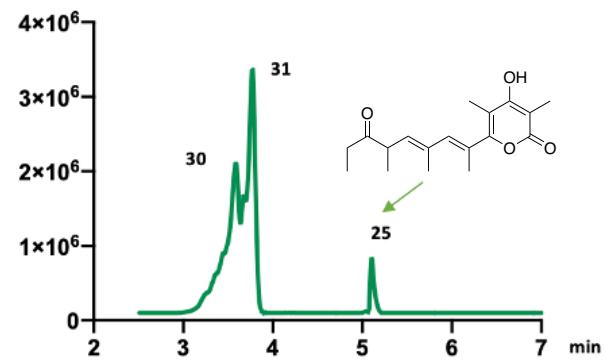
B





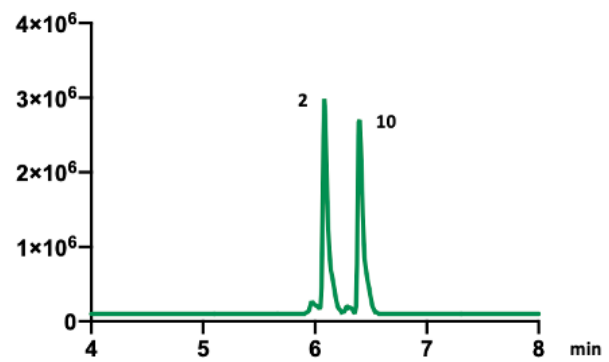
EcPKSf1-2

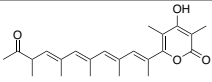
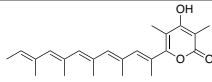
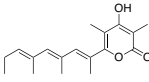
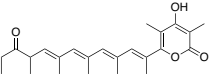
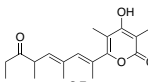
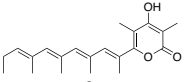
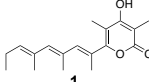
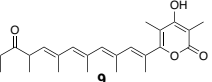
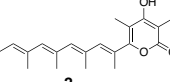
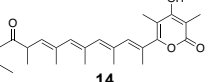
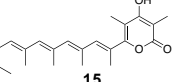
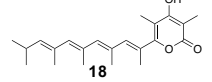
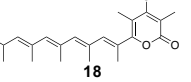
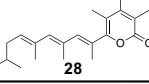
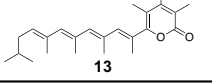
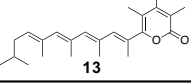
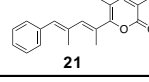
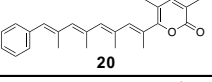
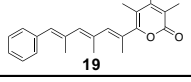
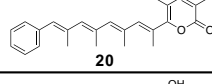
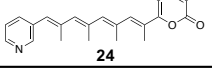
KS AT DH MT⁰ ER⁰ KR ACP



EcPKSf2-1

KS AT DH MT⁰ ER⁰ KR ACP



Starter unit	EcPKS1 ^a	EcPKS2	EcPKSf1-2 ^a	EcPKSf2-1
Acetyl-CoA	Trace 29	 16	ND	 27
Malonyl-CoA	Trace 29	ND	ND	ND
Methylmalonyl-CoA	 1	 9	 25	 2
Propionyl-CoA	 1	 9	 30	 2
Ethylmalonyl-/butyryl-CoA	ND	 14	ND	 15
Isobutyryl-CoA	ND	 18	ND	 18
Isovaleryl-CoA	 28	 13	 32	 13
Benzoyl-CoA	 21	 20	 33	 19
Benzoyl-SNAC	ND	 20	ND	ND
Nicotinyl-SNAC	ND	 24	ND	ND

^a: holoprotein activated in vitro using Sfp.