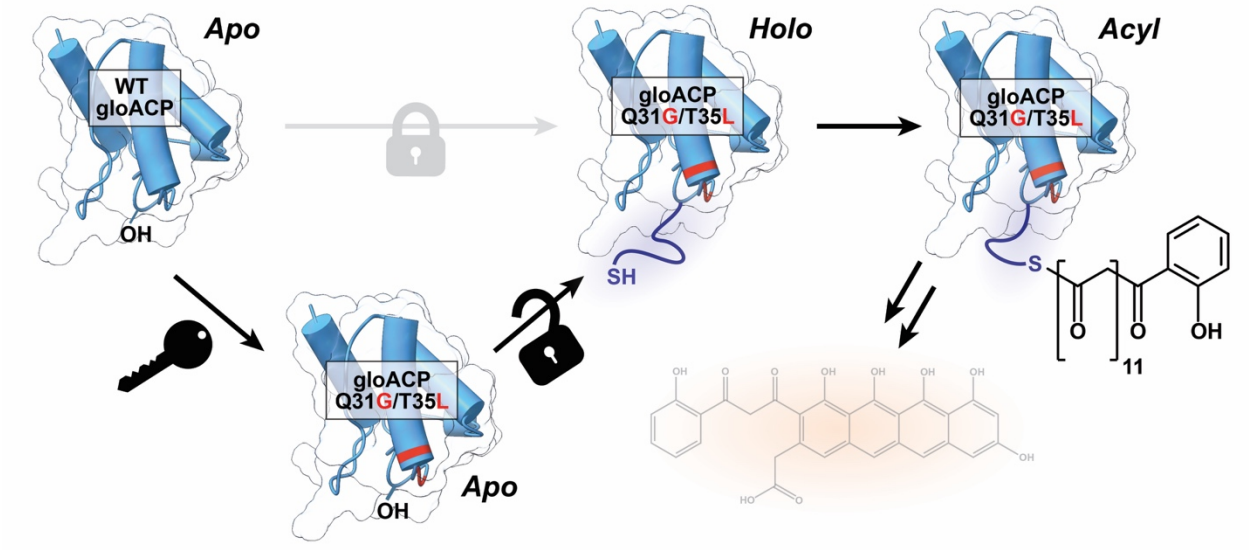
Strategic engineering unlocks *in vitro* type II polyketide biosynthesis

Kevin K. Li, Yae In Cho, Mai Anh Tran, Christoph Wiedemann, Rebecca S. Koweek, Ngọc Khánh Hoàng, Grayson S. Hamrick, Margaret A. Bowen, Bashkim Kokona, Joris Beld, Ute A. Hellmich*, and Louise K. Charkoudian*

Abstract

Microbial type II polyketides serve as powerful medicinally relevant agents. These molecules are biosynthesized by polyketide synthases (PKSs) comprised of a core ketosynthase-chain length factor (KS-CLF) and phosphopantetheinylated acyl carrier protein (*holo*-ACP). While engineering type II PKSs holds potential to unlock sustainable access to diverse bioactive molecules, the inability to obtain cognate type II KS-CLFs and *holo*-ACPs for *in vitro* studies represents a longstanding barrier. Herein, we share how the sequence and structural analysis of the *Gloeocapsa sp.* PCC 7428 ACP allowed us to tune to a requisite weak yet specific interaction with a phosphopantetheinyl transferase to afford the *holo*-ACP. This, coupled with our ability to heterologously express the cognate KS-CLF in high quantities, unlocked access to polyketide products via *in vitro* multienzyme assembly. We hope this work inspires future studies of type II PKSs that have previously evaded heterologous expression or have yet to be explored.



Main

Natural products are chemical mediators for microbial communication, competition, and evolutionary adaptation. Alongside their synthetic derivatives, these molecules account for over half of all antibacterial, immunosuppressive, and anticancer agents developed in the last four decades.[1,2] Type II polyketides stand out as a particularly prolific source of medicinally relevant natural products. Prominent examples include the anticancer agents doxorubicin and daunorubicin as well as the tetracycline class of broad-range antibiotics (*e.g.* oxytetracycline and minocycline).[3,4] These molecules are manufactured by multienzyme assemblies, termed type II polyketide synthases (PKSs), that are encoded by biosynthetic gene clusters (BGCs)

within microorganisms. The roster of discrete monofunctional enzymes includes two key players: the heterodimer ketosynthase-chain length factor (KS-CLF) and a dynamic acyl carrier protein (ACP). The ACP shuttles substrates and intermediates between their appropriate interaction sites, while the KS-CLF facilitates iterative rounds of Claisen-like condensations thereby converting acetate-based building blocks into a nascent β-ketoacyl intermediate. Other accessory type II PKS enzymes then facilitate various cyclization and oxidation events, transforming this β-ketoacyl intermediate into the programmed polyaromatic product.[5–7]

Easy access to type II PKS components would enable molecular-level investigations into the biosynthetic process as well as access to diverse molecules through biosynthetic engineering approaches.[8,9] In particular the ability of type II PKSs to transform simple monomeric building blocks into difficult-to-synthesize medicinally relevant molecules has inspired decades of research focused on the *in vitro* reconstitution of these biosynthetic macromolecular machines. Although cell-based reactions with enzymes obtained from actinomycete hosts have been used previously to produce native and engineered bioactive type II polyketides,[10–16] the relative slow growth, productivity, and difficulty for genetic manipulation of actinomycetes has severely limited the utility and progress of biosynthetic engineering work.[17,18] Attempts to heterologously express KS-CLFs from actinomycetes (a diverse phylum known to be a rich source for type II polyketides) in *E. coli* (an easily accessible and genetically tractable host), has historically been met with insoluble protein and/or low protein yields.[18] Despite intense efforts, the production of the core type PKS enzymes in a robust heterologous expression system represented a bottleneck thereby impeding the functional reconstitution of type II PKS machinery.

The recent discovery of BGCs encoding for anciently diverged type II PKSs in non-actinobacterial phyla unlocked new avenues to ameliorate the first challenge—heterologous expression of core type II PKS components for *in vitro* studies.[19–21] These newly discovered PKSs are inferred to be more closely related to paralogous *E. coli* and *B. subtilis* fatty acid synthases (FASs) than their actinobacterial PKS counterparts, and accordingly several non-actinobacterial KS-CLFs were found to express in these hosts.[22–24] Nonetheless, while the cognate non-actinobacterial ACPs could be hetereologously expressed readily in *E. coli*, the conversion from their inactive (*apo*-) state to their active (*holo*-) state proved difficult, consequently stymieing attempts to reconstitute the core type II PKS *in vitro*. Typically, ACP activation is facilitated by the covalent attachment of a coenzyme A (CoA) derived 4′-phosphopantetheine (Ppant) arm to a conserved ACP serine residue in a reaction catalyzed by a phosophopantetheinyl transferase (PPTase).[25] This post-translational modification is an essential step in polyketide biosynthesis because the Ppant arm acts as a biological tether to mediate substrate transport between ACPs and other PKS domains. While microorganisms encode a plethora of PPTases, *B. subtilis* Sfp or Sfp R4-4, a phage-selected mutant with increased substrate specificity and catalytic efficiency, are the most widely used to activate ACPs *in vitro*.[26,27] Unexpectedly, Sfp was found to be ineffective for the conversion of non-actinomycete ACPs into their *holo*-forms.[28,29] Thus, a generalizable workflow to afford *holo*-ACPs from non-actinobacterial type II PKSs has been lacking, and to our knowledge, no methods have made type II PKSs truly accessible to *in vitro* studies.[22,28–30]

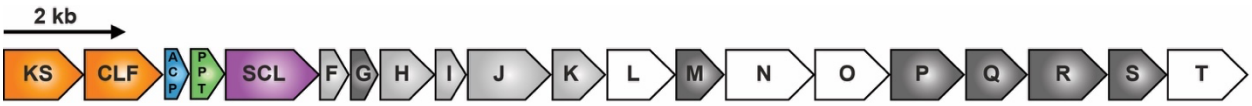
Here, through sequence alignments and structural modeling, we identified residues in non-actinobacterial type II PKS ACPs that are crucial for their *B. subtilis* Sfp-catalyzed phosphopantetheinylation. Mutation of these residues convert previously Sfp-incompatible non-actinobacterial ACPs into Sfp-compatible domains. Importantly, we also found the

corresponding KS-CLF components to be amenable to high-yield expression in *E. coli*. Together, this now allows the efficient reconstitution of all core biosynthetic enzymes of a type II PKS *in vitro*. Using this approach, we transform the non-actinomycete *Gloeocapsa sp.* PCC 7428 type II PKS ACP domain into its *holo* form and, together with the purified cognate KS-CLF, demonstrate that these PKS proteins can be used to manufacture β-ketoacyl chains that can be folded into polyketide structures. Thus, our findings provide the basis for the streamlined production of type II polyketides through the strategic engineering and *in vitro* reconstitution of non-actinomycete type II PKSs. We hope that lessons learned from these studies will inspire future investigations of type II PKSs from diverse phyla, including those that have evaded heterologous expression or have yet to be explored.

Results

Despite successful heterologous expression and purification of a type II PKS, *in vitro* activation of its ACP proves challenging.

To assess whether non-actinobacterial type II PKSs are in principle amenable to heterologous expression in *E. coli* and to production of polyketide compounds *in vitro*, we turned to the cyanobacterium *Gloeocapsa* sp. PCC 7428. It contains a diverse array of natural product BGCs, including a single BGC predicted to encode for a type II PKS that produces polyaromatic molecules—which we named gloPKS. The gloPKS cluster consists of typical type II PKS core enzymes (gloACP and the heterodimeric gloKS-CLF) in addition to various priming, accessory, and tailoring domains (Fig. 1). The genes encoding for gloKS-CLF were amplified from the organism's genomic DNA. Given that transcriptional coupling between partner KS and CLF genes has previously shown importance for stability, solubility, and complex formation of these heterodimers in heterologous hosts,[28] the native transcript overlap between *gloKS* and *gloCLF* was maintained. This resulted in the successful heterologous co-expression of the two proteins under a single promoter and high-yield purification (>20 mg/L) from *E. coli* BL21(DE3) cells (Supplementary Fig. S1).



Protein	Deduced Function	Protein	Deduced Function	Protein	Deduced Function	Protein	Deduced Function
gloKS*	Ketosynthase	gloF	Cyclase	gloK	Thioesterase	gloP	Glycosyltransferase
gloCLF*	Chain-length factor	gloG	Carboxymuconolactone decarboxylase	gloL	Unknown	gloQ	Glycosyltransferase
gloACP*	Acyl carrier protein	gloH	Cyclase	gloM	Glycosyltransferase	gloR	Glycosyltransferase
gloPPT*	Phosphopantetheinyl transferase	gloI	Cyclase	gloN	Unknown	gloS	Glycosyltransferase
gloSCL*	Salicylate CoA ligase	gloJ	Salicylate synthase	gloO	Unknown	gloT	Unknown

Figure 1. *Gloeocapsa* sp. PCC 7428 type II PKS BGC organization. Deduced gene functions for core PKS enzymes (asterisk), accessory PKS enzymes (light gray), tailoring enzymes (dark gray), and domains of unknown function (white).

In addition, the cognate ACP domain, gloACP, was heterologously expressed and successfully purified in its *apo*-form (Supplementary Fig. S1-2). Attempts were made to activate the ACP into its *holo*-form. However, the coexpression and copurification of gloACP with gloPPT, the native gloPKS BGC PPTase, in *E. coli*, as well as the individual expression and subsequent co-incubation of purified gloPPT and gloACP did not convert the ACP to the desired *holo* product as seen by mass spectrometry (Supplementary Fig. S3-5). Furthermore, we observed minimal expression of gloPPT in *E. coli*, both with gloACP or individually (Supplementary Fig. S6).

It has been previously shown that putative CoA:ligase domains promote the ability of non-actinobacterial in-cluster PPTases to catalyze phosphopantetheinylation, although the mechanism behind this highly synergistic effect remains unknown.[22,30] Therefore, we also heterologously expressed and purified gloSCL, a salicylate CoA:ligase domain from the gloPKS, (Supplementary Fig. S1). This protein was co-incubated with gloACP alongside its co-factors (ATP and salicylic acid) to facilitate the phosphopantetheinylation reaction. However, this approach also did not lead to the production of *holo*-gloACP (Supplementary Fig. S7). Thus, we conclude that an as of yet unidentified pathway is responsible for the conversion of *apo*- to *holo*-gloACP in *Gloeocapsa*.

As additional routes to obtain *holo*-gloACP, we furthermore explored the possibility of using *B. subtilis* Sfp, a well-established PPTase for carrier proteins (CPs) by (i) expressing and purifying gloACP from *E. coli* BAP1 cells, a genetically engineered derivative of the BL21 (DE3) strain carrying background expression of Sfp[31] and (ii) co-incubating recombinant Sfp R4-4[26] with purified gloACP. However, again both approaches did not yield a phosphopantetheinylated ACP, as assessed by mass spectrometry (Supplementary Fig. S8).

This shows that type II PKS ACP activation *in vitro* is not straightforward, even when the type II PKS components and activation enzymes can be expressed heterologously. We concluded that in the case of the native PPTase, a hitherto unknown factor may play a role. Furthermore, incompatibilities between gloACP and Sfp R4-4 could prevent gloACP activation via this conventionally used modification enzyme.

Sfp R4-4 binds to gloACP but does not facilitate gloACP phosphopantetheinylation.
To determine if the observed lack of conversion from *apo*- to *holo*-gloACP is due to structural differences between gloACP and ACPs compatible with Sfp R4-4, NMR spectroscopy was used. Upon determining the backbone and sidechain NMR assignments of ^{13}C, ^{15}N-labeled gloACP (Fig. S9), we calculated a chemical shift derived structural model using CS-Rosetta[32] (Fig. 2A, Supplementary Fig. S10). The CS-Rosetta model of gloACP was then compared with available X-ray and NMR structures of ACPs from canonical type II FAS (*E. coli* AcpP, 1T8K), actinobacterial type II PKS (*Streptomyces coelicolor* ActACP, 2K0Y), and type I PKS (*Saccharopolyspora erythraea*, DEBS ACP2, 2JU2) systems (Fig. 2B). GloACP differs from AcpP, and ActACP by two short helical insertions between helix I and helix II, which we accordingly named helix I′ and helix II′. Furthermore, gloACP lacks helix III', an insertion between helix II and III seen in other ACPs. Despite this, the overall structures of the different ACP domains and the position of the conserved site serine residue are highly similar, indicating

that structural differences are unlikely to be the cause for the inability of Sfp R4-4 to activate gloACP.

Next, we tested whether Sfp R4-4 is able to interact with gloACP by titrating ^{15}N-labeled 11 kDa gloACP with unlabeled 27 kDa Sfp R4-4 in a chemical shift perturbation assay (Fig. 2C). Already at sub-stochiometric ratios, severe line broadening was observed, suggesting the formation of a large molecular weight complex. At a 1:1 ratio, all signals of the gloACP disappeared, arguing for a tight complex formation and indicating that the inability of Sfp R4-4 to post-translationally modify gloACP is not due to a lack of protein interaction.

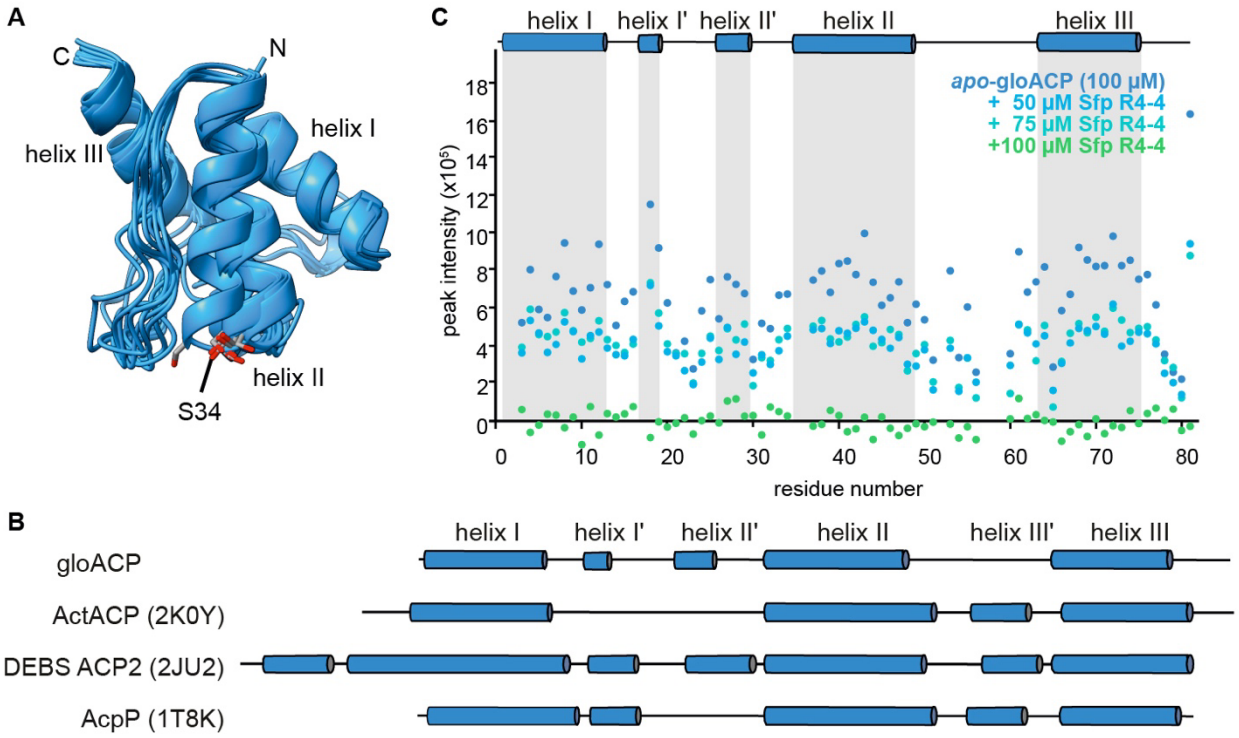


Figure 2. gloACP interacts tightly with Sfp R4-4. (A) The ten lowest energy structural models of WT gloACP obtained by CS-Rosetta. (B) Topology models of WT gloACP in comparison to the structures of prototypical ACP domains: a type II PKS ACP (ActACP, PDB: 2K0Y) from *Streptomyces coelicolor*, a type I PKS ACP (DEBS ACP2, PDB: 2JU2) from *Saccharopolyspora erythraea*, and a type II FAS ACP (AcpP, PDB: 1T8K) from *Escherichia coli* (cylinders represent α-helical secondary structure). (C) NMR peak intensities of ^{15}N-labeled gloACP upon Sfp R4-4 titration. Severe line broadening at a 1:1 molar ratio indicates tight complex formation.

Two non-polar residues are crucial to enable gloACP phosphopantetheinylation by Sfp

To elucidate the molecular basis of the inability of Sfp to phosphopantetheinylate gloACP and possibly ameliorate it, we carried out a multiple sequence alignment between gloACP and CP domains known to be modified by Sfp (Fig. 3A, Supplementary Fig. S11). Recently, the structure of *Bacillus brevis* tyrocidine A synthetase peptidyl carrier protein (TycC3_PCP) domain in complex with Sfp revealed that PCP residues G42, L46, and M49 are critical for facilitating phosphopantetheine transfer.[33] Other Sfp-modifiable ACPs also carry a glycine at the analogous position 42, and non-polar residues at the analogous positions 46 and 49. In contrast, gloACP contains polar residues at analogous locations to positions 42 and 46 of TycC3_PCP (Q31 and

T35 highlighted in magenta, Fig. 3A). To investigate if these two sequence variations within helix II of the ACP are sufficient to explain the difference in phosphopantetheinylation behavior between gloACP and the other CPs, we generated single (gloACP Q31G and gloACP T35L) and double gloACP mutants (gloACP Q31G/T35L and gloACP Q31G/T35I). Using an MS-based detection assay, the level of phosphopantetheinylation for each heterologously expressed and purified mutant upon incubation with Sfp R4-4 was assessed (Supplementary Fig. S12–16). While the single gloACP Q31G mutant was not modified, single and double mutants gloACP T35L, Q31G/T35L, and Q31G/T35I showed conversion to *holo*-ACPs in varying degrees. This indicates the presence of a nonpolar residue at position 35 in gloACP is most crucial in enabling this protein's Sfp-catalyzed phosphopantetheinylation. The additional modification of residue 31 supplements this process, as GloACP Q31G/T35L showed the most efficient phosphopantetheinylation of all our gloACP mutants.

Accordingly, the interaction of GloACP Q31G/T35L with Sfp R4-4 was further investigated by NMR spectroscopy for further molecular insight (Fig. 3B, C). The NMR backbone assignment of WT gloACP could be transferred to the mutant to a substantial degree, indicating a large structural similarity (Supplementary Fig. S9). The most notable chemical shift differences between the two proteins are located in helix II and the loop between helix II and helix III (Fig. 3B). Interestingly, when unlabeled Sfp R4-4 was titrated with isotope-labeled gloACP Q31G/T35L, resonances were still visible at an equimolar ratio (Fig. 3C), indicating a weakening of the interaction with Sfp R4-4 upon mutation of gloACP. This decrease in binding affinity may enable the formation of a loose protein interface enabling efficient Ppant arm transfer. Encouragingly, these results indicate that we may have discovered a generalizable route to imbue previously non-Sfp competent ACP domains with the ability to be efficiently phosphopantetheinylated by Sfp, a crucial step for the *in vitro* reconstitution of non-actinobacterial type II PKS activity.

Lessons learned from gloACP engineering enables activation of other non-actinobacterial ACPs

To test whether our strategy in modifying the ACP/Sfp interface through simple ACP helix II point mutations is broadly applicable in converting Sfp-incompatible ACPs into compatible modules, two additional ACPs were selected based on their helix II sequences for study: type II PKS ACPs from *Delftia acidovorans* (dacACP) and *Paenibacillus sp.* FSL R7-277 (panACP) (Fig. 3A). Both dacACP and panACP contain a threonine at positions analogous to gloACP 35/ TycC3_PCP 46. Thus, based on our prior work with gloACP, we believed both ACPs to be incompatible with Sfp. PanACP contains 2 additional residues predicted to cause Sfp-incompatibility (analogous positions to gloACP 31 and 38/ TycC3_PCP 42 and 49), while dacACP contains residues at these locations predicted to be compatible for Sfp-interaction.[33] Following phosphopantetheinylation assays with Sfp R4-4, WT panACP displayed no phosphopantetheinylation, while WT dacACP displayed very minimal levels (Supplementary Fig. S17-20). This further suggests the importance of a nonpolar residue analogous to gloACP 35/ TycC3_PCP 46 in facilitating the Sfp-ACP interactions necessary for efficient phosphopantetheinylation.

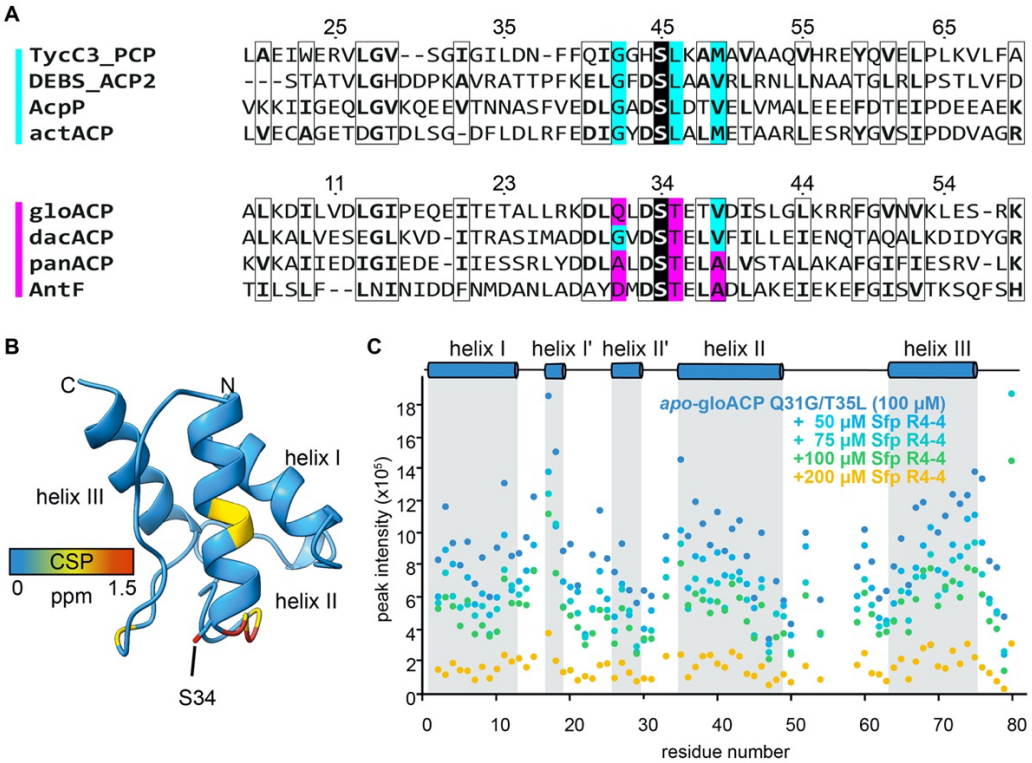


Figure 3. An ACP double mutant is efficiently phosphopantetheinylated by Sfp R4-4 through formation of a loose complex. (A) Multiple sequence alignment of various CPs known to be modified by Sfp (top four with cyan bar) and non-actinobacterial ACPs that demonstrate Sfp-incompatibility (bottom four with magenta bar). GloACP, dacACP, and panACP were used in this study, while AntF is a previously characterized non-actinobacterial ACP from *Photorhabdus luminescens* TT01 that is known to lack compatibility with Sfp.[29] The serine point of attachment of the phosphopantetheine arm is highlighted in black for each CP sequence. Black boxes represent residues containing greater than 70% physicochemical similarity. Residues highlighted in cyan are predicted to confer Sfp-compatibility, while residues highlighted in magenta are predicted to result in increased Sfp-ACP incompatibility. (B) Chemical shift perturbation (CSP) between gloACP and gloACP Q31G/T35L (Supplementary Fig. S9) mapped on the CS-Rosetta model of WT gloACP. (C) Peak intensity of ^{15}N-labeled gloACP Q31G/T35L (100 μM) upon titration with unlabeled Sfp R4-4.

To further investigate this connection, we followed our previously established workflow, creating the mutants dacACP T43L and panACP A30G/T34L/A37V to again emulate the analogous residues found in our newly Sfp-compatible gloACP Q31G/T35L domain. The dacACP and panACP mutants were both readily phosphopantetheinylated by Sfp R4-4 *in vitro* (Supplementary Fig. S21-24). Together these results show that 1) ACP sequence can be used to predict Sfp R4-4 compatibility and 2) the strategic modification of residues at positions analogous to gloACP 31, 35, 38/ TycC3_PCP 42, 46, 49 represents a generalizable strategy to enable the activation of type II PKS ACPs as a prerequisite for *in vitro* polyketide production.

In vitro polyketide production by a reconstituted minimal type II PKS

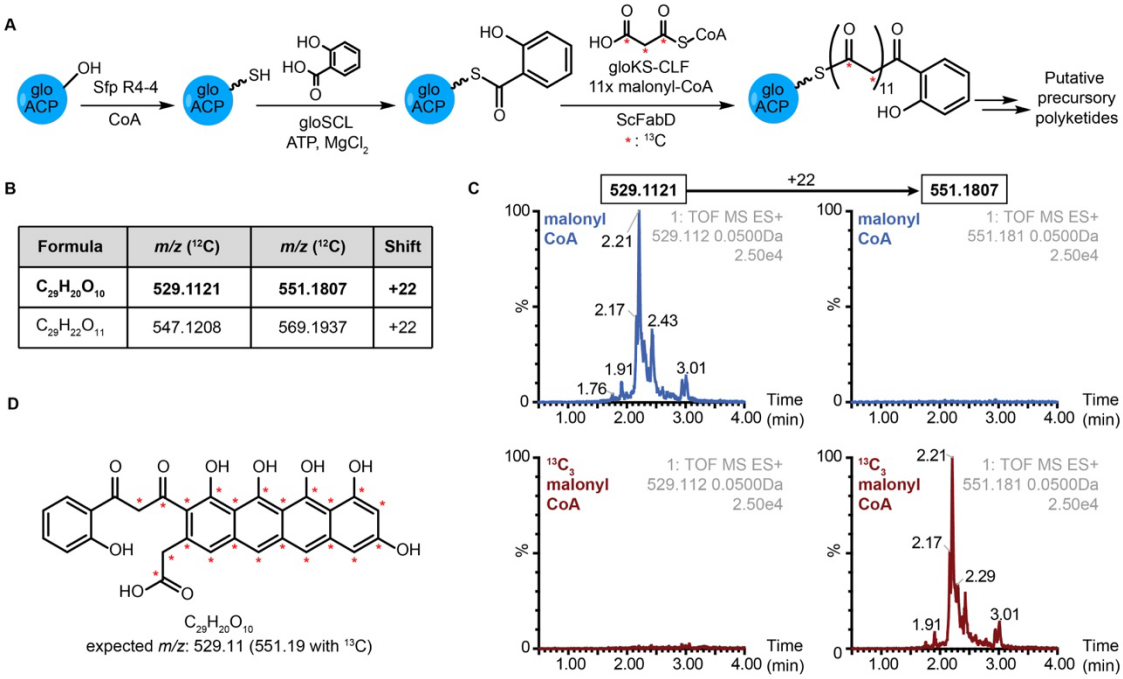
Next, we sought to determine whether the mutated ACP was able to engage in the decarboxylative Claisen-like condensations required to build a type II polyketide. We took advantage that the gloPKS system shows high sequence and BGC organization similarity to the previously characterized thermorubin-producing PKS from *Laceyella sacchari* (Supplementary

Table S1).[34] Thermorubin is a tetracyclic antibiotic comprising a 29-carbon polyketide backbone with seven carbons derived from salicylate and 22 carbons derived from acetate-based malonyl-CoA. It has been suggested that the carbon backbone is initially formed as a protein-tethered salicyl-primed dodecaketide intermediate by the core biosynthetic domains—specifically, via salicylate:CoA ligase mediated priming and iterative rounds of ACP-KS-CLF directed extension.[35] Afterwards, offloading and several uncharacterized ring closure events transform the nascent β-ketoacyl chain into precursors that are ultimately converted into a final tetracyclic structure through tailoring enzyme-mediated modifications.[35] Given the resemblances between the two systems, we anticipated that the gloPKS would use similar chemical substrates as the thermorubin PKS, and execute analogous biosynthetic reactions leading to the formation of comparably composed polyaromatic compounds (Fig. 4A).

Using Sfp-modified *holo*-gloACP Q31G/T35L, we first performed *in vitro* reactions using the inferred native priming enzyme (gloSCL) and starter unit (sodium salicylate), as well as the necessary cofactors (ATP and $MgCl_2$). Liquid chromatography mass spectrometry (LC-MS) revealed the complete conversion of *holo*-gloACP Q31G/T35L to the salicyl-form thus demonstrating that mutagenesis of the ACP for phosphopantetheinylation does not interfere with its ability to interact with substrates and priming enzymes (Supplementary Fig. S25). Next, we examined the ACP domain's malonylating activity, a critical step in the initiation of polyketide extension. The majority of ACPs studied to date require malonyl-CoA:acyl carrier protein transacylases (MCATs) to catalyze this process.[36,37] However, in a small number of cases, MCATs are not required due to the ability of the ACP itself to catalyze this process in what is referred to as self-malonylation.[38] Interestingly, *holo*-gloACP Q31G/T35L demonstrated some self-malonylating activity, becoming partially malonylated upon incubation with malonyl-CoA (Supplementary Fig. S26). Co-incubation with MCATs from either *Streptomyces coelicolor* (ScFabD) or *E. coli* (EcFabD) resulted in the full conversion from *holo-* to malonyl-gloACP Q31G/T35L (Supplementary Fig. S27 and S28), furthermore highlighting the compatibility of a mutated ACP to interact with a variety of binding partners from various sources and retaining its requisite functionality of loading its putative salicyl priming unit and malonyl building blocks.

Next, we tested whether *holo*-gloACP Q31G/T35L was still capable to partner with its cognate KS-CLF thereby facilitating the Claisen-like bond formation required to build poly-β-ketoacyl intermediates. All core biosynthetic gloPKS components (*holo*-gloACP Q31G/T35L, gloKS-CLF, and gloSCL) were purified from *E. coli*, co-reconstituted and incubated with their inferred substrates (salicylic acid, malonyl CoA) and cofactors (ATP, $MgCl_2$). This resulted in the production of a set of putative polyaromatic polyketides, as shown by LC-MS (Fig. 4B, C). As hypothesized based on the biosynthetic logic of thermorubin production, we observed numerous isomers of two mass peaks (*m/z* 529 and 547 $[M+H]^+$) that corresponded to the expected condensation of 11 malonyl-CoA units and priming of one salicylate unit. High-definition mass spectrometry identified the molecular formulas represented by these peaks as $C_{29}H_{20}O_{10}$ and $C_{29}H_{22}O_{11}$, respectively, matching the expected composition of the polyketide products. To confirm that the observed peaks were due to malonyl-CoA condensation reactions facilitated by the core PKS enzymes, we repeated the reaction with $^{13}C_3$-labeled malonyl-CoA, observing a shift of 22-amu for both sets of peaks (Fig. 4B, Supplementary Fig. S29). Further MS/MS analyses indicated that the identified structures produced by our reconstituted minimal gloPKS are consistent with previously hypothesized untailored thermorubin precursors (Figure 4D, Supplementary Fig. S30). Taken together, these findings show that *holo*-gloACP Q31G/T35L remains a functional unit within the gloPKS and confirms that the strategic mutagenesis

approach of modifying non-actinobacterial ACP domains opens up these systems for *in vitro* investigations.



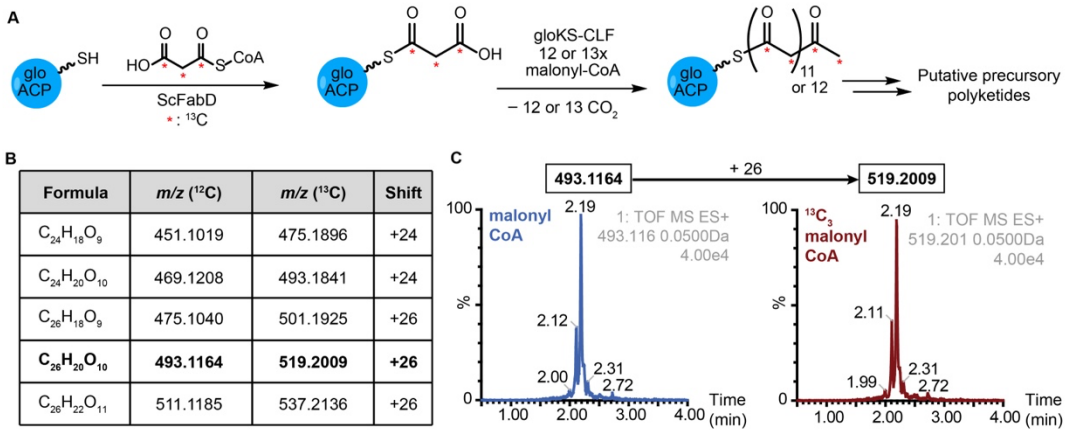
Formula	*m/z* (^{12}C)	*m/z* (^{12}C)	Shift
$C_{29}H_{20}O_{10}$	529.1121	551.1807	+22
$C_{29}H_{22}O_{11}$	547.1208	569.1937	+22

Figure 4. *In vitro* polyketide production via co-reconstitution of mutated gloACP core type II *Gloeocapsa* PKS components. (A) Proposed *in vitro* gloPKS polyketide production scheme. Red stars represent carbons derived from malonyl-CoA and anticipated to be labeled during $^{13}C_3$-malonyl-CoA isotope experiments. (B) Chemical formulas and experimental *m/z* values (^{12}C and ^{13}C) of putative precursor polyketide metabolites obtained from LC-MS experiments. (C) Extracted ion chromatograms of the putative metabolite $C_{29}H_{20}O_{10}$. Upon $^{13}C_3$-malonyl-CoA supplementation, a mass shift of +22 amu is observed. Additional extracted ion chromatograms of putative precursor polyketide metabolites can be found in Supplementary Fig. S29. (D) Proposed chemical structure of the putative metabolite $C_{29}H_{20}O_{10}$ produced by the reconstituted type II gloPKS. Red stars represent carbons display a mass shift upon $^{13}C_3$-malonyl-CoA labeling.

Priming of *holo*-gloACP Q31G/T35L with malonyl-CoA leads to the production of alternative molecular outputs

While salicylate is the inferred priming unit of the *Gloeocapsa* type II PKS, we observed evidence that many polyketide products appeared to be derived solely from malonyl-CoA even when salicylic acid was present (Fig. 5A). These polyketides (with *m/z* of 451, 469, 475, 493, and 511 $[M+H]^+$) showed isotopic shifts of 24-amu (451, 469) and 26-amu (475, 493, 511) upon ^{13}C-malonyl CoA feeding, consistent with dodecaketides and tridecaketides (Fig. 5B). Through hgh-definition MS, the molecular formulas of $C_{24}H_{18}O_9$ (451), $C_{24}H_{20}O_{10}$ (469), $C_{26}H_{18}O_9$ (475), $C_{26}H_{20}O_{10}$ (493), and $C_{26}H_{22}O_{11}$ (511) could be assigned, suggesting these polyketides are produced solely from the condensation of 12 and 13 molecules of malonyl-CoA, respectively. These unexpected findings prompted us to compare the priming specificity of the *Gloeocapsa* type II PKS to other type II PKSs that integrate non-acetate (*e.g.* benzoate, anthranilate, propionate) starter units into their programmed products.[39–41] In the absence of gloSCL/ salicylate and presence of malonyl-CoA, *holo*-gloACP Q31G/T35L and gloKS-CLF solely produced the previously identified molecules of *m/z* 451, 469, 475, 493, and 511, confirming these polyketides to be entirely malonyl-CoA derived (Fig. 5C and Supplementary Fig. S31).

This demonstrates that gloPKS reactions with the *holo*-gloACP Q31G/T35L can manufacture polyketides in the absence of its putative native priming component, a finding consistent with prior studies that demonstrate type II PKSs containing initiation modules can have their core biosynthetic rosters shuffled to alter the chain lengths of polyketides produced.[42–44]



Formula	*m/z* (^{12}C)	*m/z* (^{13}C)	Shift
$C_{24}H_{18}O_9$	451.1019	475.1896	+24
$C_{24}H_{20}O_{10}$	469.1208	493.1841	+24
$C_{26}H_{18}O_9$	475.1040	501.1925	+26
$C_{26}H_{20}O_{10}$	**493.1164**	**519.2009**	**+26**
$C_{26}H_{22}O_{11}$	511.1185	537.2136	+26

Figure 5. Type II PKS cab be primed with non-cognate starter unit to lead to an alternative product range. (A) Proposed *in vitro* reaction scheme of gloPKS polyketide production in the absence of gloSCL and salicylic acid. (B) Chemical formulas and the experimental *m/z* values (^{12}C and ^{13}C) of putative precursor polyketide metabolites produced by the reconstituted gloPKS in the absence of gloSCL. (C) Extracted ion chromatograms of the putative metabolite $C_{26}H_{20}O_{10}$. $^{13}C_3$-malonyl-CoA supplementation leads to a mass shift of +26 amu in both cases. Additional extracted ion chromatograms of putative precursor polyketides can be found in Supplementary Fig. S31.

Discussion

Decades of bioprospecting research has been fueled by the strong clinical track-record of microbial type II polyketides and the promise of biosynthesis to open up sustainable routes to these natural products. Strategic engineering efforts of type II PKSs should thus enable access to structurally complex 'unnatural' natural products.[10,11,14] However, the inability to obtain high quantities of core type II PKS enzymes via heterologous expression has stymied progress and is likely a key reason why type II PKS engineering efforts lag behind their type I counterparts.[18,31] Furthermore, the inability of Sfp, the archetypical PPTase for *in vitro* applications, to phosphopantetheinylate numerous type II PKS ACP domains[28,29] creates an additional hurdle in the *in vitro* reconstitution of these biosynthetic complexes. Herein we demonstrate that it is possible to obtain active type II PKS core components in high quantities via heterologous expression in *E. coli* through the strategic selection of KS-CLF genes and generalizable point mutation of their ACPs.

We show that the non-actinobacterial *Gloeocapsa* KS-CLF can be expressed and purified from heterologous expression in *E. coli*. In addition to allowing to skip the severely more demanding growth and purification protocols from the native hosts, the obtained production yields from *E. coli* of ~25 mg/L are roughly 25-fold greater than what is typically obtained from actinobacterial hosts.[45,46] We believe this manageability of the gloPKS to be due to (i) its evolutionary roots, inferred to be more closely related to *E. coli* than its actinomycete counterparts, and (ii) the fact that the *ks* and *clf* genes are natively transcriptionally coupled.[19,22,28]

The observation that the gloACP was not activated to its *holo*-form via the promiscuous *B. subtilis* Sfp R4-4 led us to explore their interaction via sequence and structural analyses. Multiple sequence alignment of Sfp incompatible ACPs to Sfp compatible ACPs, and subsequent comparison to the available TycC3_PCP-Sfp structure[33] informed our strategic design of the confirmed Sfp R4-4 compatible gloACP Q31G/T35L mutant. In contrast to what might be expected, decreasing rather than increasing the gloACP-Sfp R4-4 binding affinity was a key driver in conferring the functional interaction between the two proteins. Thus, the inability of Sfp to phosphopantetheinylate non-actinobacterial ACPs is likely not due to a lack of interaction, but an overly strained binding that ultimately inhibits the promiscuous PPTase's ability to catalyze phosphopantetheine transfer. These results could provide key insight for the gain of phosphopantetheinylation viability of previously inaccessible CPs from other natural product fields and could be applied to other type II synthases' protein partnerships as well. Further, this observation could guide the development of potential peptidyl inhibitors for the targeting of bacterial PPTases implicated in common pathogens' associated virulence factors.[47,48]

The ability to access gloPKS core biosynthetic enzymes for *in vitro* investigations allowed us to explore the substrate specificity and product promiscuity of the system. GloACP Q31/T35 could be loaded with a salicyl or malonyl group. This led to the production of dodecaketides and tridecaketides (malonyl only) as well as dodecaketide thermorubin-like products (malonyl and salicyl; Figure 4D).[35] The observed product diversity can be attributed to how the CLF domain directs the chain lengths of polyketides produced by type II PKSs,[49] with the identity of the polyketide starter unit directing this mechanism.[42–44] It is enticing to assume that the gloCLF cavity may naturally accommodate large hydrophobic moieties. Therefore, in the absence of a salicylate group on its nascent β-ketoacyl chain, the addition of multiple keto groups to the gloCLF-tethered growing intermediate could allow the growth of large, and structurally complex aromatic polyketides. It furthermore implies that the incorporation of unnatural starter units in type II PKS bioengineering could be used to dictate chain lengths of novel 'unnatural' polyketides produced.

Taken together, these data suggest that the gloPKS system is capable of producing a variety of polyketides and that its product profile can be altered by changing the roster of core enzymes present. This biosynthetic versatility is possibly further enhanced through the self-malonylation capability of the gloACP Q31G/T35L mutant and/or incorporating unnatural priming units. Given that the strategic mutation workflow for converting Sfp-incompatible ACPs to Sfp-compatible ACPs was extended to other systems for which the KS-CLFs can be obtained in high yield via heterologous expression in *E. coli*, the impact of this work extends well beyond the *Gloeocapsa* system. By applying the lessons learned from this study to diverse type II PKS systems, we can unlock the playbook for a range of type II PKS systems and ground rules for individual components, thereby opening the way to realizing the full promise of these remarkable bacterial assembly lines.

Methods

Native Organism Culturing and DNA Isolation.

Gloeocapsa sp. PCC 7428 (ATCC® 29159™) was purchased as a liquid culture from the American Type Culture Collection (ATCC®) and inoculated in 30 mL of BG-11 liquid medium (ATCC® medium 616). Cultures were grown at room temperature with indirect light exposure for

6-7 weeks with aeration and no shaking. Genomic DNA was isolated following the protocol described in Saha *et al*.[50]

Molecular cloning.
E. coli DH5α competent cells were used for all cloning experiments. AntiSMASH 5.0 was used to identify genes corresponding to *Gloeocapsa sp.* PCC 7428 (ID: CP003646) ACP, KS-CLF, PPT, and SCL genes.[51] Each gene was amplified from purified genomic DNA using primers listed in Supplementary Table S2 and standard Q5 High-fidelity DNA polymerase. For genes corresponding to ACPs from *Paenibacillus* sp. FSL R7-277 and *Delftia acidovorans,* codon optimized DNA was purchased from Twist bioscience and amplified following similar procedures. All thermocycling conditions are listed in Supplementary Table S3. Gel-purified amplimers were inserted into linearized pET28a vector backbone between NdeI and EcoRI via Gibson assembly[52] using an NEB Q5 high-fidelity 2X master mix kit, generating respective plasmids listed in Supplementary Table S2. The expression plasmid corresponding to EcFabD was provided as a gift from the Campopiano Research group at University of Edinburgh and the plasmid corresponding to ScFabD was purchased from Gene Universal.

Multiple sequence alignment (MSA).
Amino acid sequences of all carrier proteins were aligned using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and ESPript 3.0 (https://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi).[53]

Site directed mutagenesis.
NEB Q5 site directed mutagenesis kit was used for all amplification and mutagenesis work. The pJH1 vector was amplified with four distinct primer sets to create the gloACP mutants Q31G/T35L, Q31G/T35I, Q31G, and T35L (Supplementary Table S2). pPaACP and pDaACP vectors were amplified using similar conditions to create mutants panACP A30G/T34L/A37V and dacACP T43L. All amino acid sequences from each plasmid (confirmed by sequencing) and the primers are provided in Supplementary Table S2.

Protein expression & purification.
All proteins were expressed and purified using similar procedures, which is described in the Detailed Methods section in Supplementary information.

NMR spectroscopy.
Uniformly ^{15}N and $^{13}C/^{15}N$ labeled N-terminal His_6-tagged gloACP were expressed in *E. coli* BL21(DE3) gold cells grown in isotope (1 g/L $^{15}NH_4Cl$ and 2 g/L U-^{13}C-glucose as sole nitrogen and carbon source, respectively) supplemented M9 medium using a spin-down protocol.[54] First, transformed *E. coli* cells were grown in LB medium to an OD_{600} = 0.4–0.6 and then centrifuged (25 °C, 5000 x *g*, 10 min). The cell pellets were resuspended in M9 medium supplemented with 50 *μ*g/ml kanamycin (Kan) and 312 *μ*M IPTG. Cells were harvested via centrifugation. Cell pellets were resuspended in 50 mM NaP_i pH 7.6, 300 mM NaCl, 10 mM imidazole, 10% glycerol, protease inhibitor mix (1:1000 dilution) a spatula tip of lysozyme, DNase, RNase and cells were disrupted by sonication. Cell debris was removed by centrifugation (10.000 x g, 30 min, 4 °C) and the supernatant was mixed with Ni-NTA beads for 1 h on an end-over-end rotator. The mixture was transferred to a gravity column and was washed with 10 CV 50 mM NaPi pH 7.6, 100 mM NaCl, 10 mM imidazole, 10% glycerol, and 10 CV 50 mM NaPi pH 7.6, 100 mM NaCl, 30 mM imidazole, 10% glycerol. Proteins were eluted stepwise with 50 mM NaPi pH 7.6, 100 mM NaCl, 250 mM imidazole, 10% glycerol. A BCA assay was performed to

determine fractions that contained proteins. These fractions were pooled together and concentrated. Purification via SEC using HiLoad prep grade 16 / 60 Superdex 75 column and 50 mM NaP_i pH 7.6, 100 mM NaCl was carried out. All NMR experiments were performed at 298 K on Bruker AVANCE III HD 600 MHz spectrometers equipped with cryogenic triple resonance probes (Bruker, Rheinstetten). A set of triple resonance assignment experiments (HNCO, HN(CA)CO, HNCA and HNCACB) implemented in the standard Bruker TopSpin pulse sequences library were recorded for both wild-type gloACP and gloACP Q31G/T35L at a concentration of 300–410 μM in 50 mM NaP_i pH 7.5, 100 mM NaCl, , 5% d_6-glycerol. The samples were supplemented with 10% D_2O for field-frequency locking and 0.1 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS). The ^1H signal of DSS was set to 0 ppm for direct referencing and the ^{13}C and ^{15}N chemical shifts were indirectly referenced according their magnetogyric ratios.[55] All spectra were processed with Bruker TopSpin™ 3.2 or 4.1. Data analysis of WT gloACP, gloACP Q31G/T35L, and *holo*-gloACP (Q31G/T35L) was carried out in CCPN V3.1.[56]

For titration experiments with Sfp R4-4, ^{15}N-labeled WT gloACP and gloACP Q31G/T35L were used. Spectra for both proteins were recorded at a concentration of 100 μM. Unlabeled Sfp R4-4 were added in a concentration range from 50–200 μM.

CS-Rosetta modelling.
Three-dimensional gloACP structural models were calculated with CS-Rosetta[32] (version 2.01rev2019.06 using 40 CPUs within the NMRbox virtual environment)[57] using experimental backbone ^{13}CO, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^1H^{\alpha}$, $^1H^N$ and $^{15}N^H$ chemical shift restraints. Before setting up the CS-Rosetta run, the referenced gloACP backbone chemical shifts were submitted to TALOS-N[58] for generation of dihedral restrains and predicting highly flexible or disordered regions. Since TALOS-N predicted highly dynamic amino acid stretches at the N- and C-termini of the gloACP construct the N-terminal purification tag (residue -19-0) and the last five C-terminal residues (76-80) were removed for the modeling procedure. To facilitate proper folding, repulsive energy was used for those residues during structure calculation. gloACP residues 54-58 were included in the calculation run although classified as dynamic by TALOS-N.

The CS-Rosetta run was considered to have converged based on the backbone C^{α}-atom RMSD (root mean square deviation) < 2 Å between the lowest re-scored energy structures. 1292 out of the 10,000 calculated structural models have a RMSD < 2 Å to the lowest energy model (Fig. S9). The 10 lowest re-scored energy models with a mean C^{α}-RMSD of 0.84+/-0.33 Å were chosen.

Analysis of proteins by LC-MS and tandem proteolysis mass spectrometry.
Proteins were analyzed by LC-MS (Supplementary Table S4 and Fig. S2-5, 7, 8, 12-28), and tandem proteolysis (Supplementary Fig. S32). For LC-MS, protein samples were diluted to ~10 μM using HPLC grade water and analyzed using an Agilent Technologies InfinityLab G6125B LC/MS coupled with an Agilent 1260 Infinity II LC system loaded with a Waters XBridge Protein BEH C4 reverse phase column (300A, 3.55 μM, 2.1 mm x 50 mm). Samples were heated to 45 °C and analyzed via electrospray ionization mass spectrometry (ESI-MS) in the positive mode. The following gradient was used: 0–1 min 5% B; 1–3.1 min 95% B; 3.1–4.52 min 95% B; 4.52–4.92 min 5% B; 4.92–9 min 5% B (solvent A = water + 0.1% (v/v) formic acid; solvent B = acetonitrile + 0.1% (v/v) formic acid). Following data collection, mass spectra were

deconvoluted using ESIprot online[59] and plotted using Origin (Version 8.60. OriginLab Corporation, Northampton, MA).

Tandem proteolysis was performed on purified gloKS-CLF by the Proteomics and Metabolomics Facility of the Wistar Institute located in Philadelphia, PA. Samples were reduced with tris (2-carboxyethyl) phosphine (TCEP), alkylated with iodoacetamide, and digested with trypsin. Resulting tryptic peptides were injected into a UPLC Symmetry trap column (180 m i.d. × 2 cm packed with 5 μm C18 resin; Waters) and separated using their standard 90-minute gradient formed by solvent A (water + 0.1% (v/v) formic acid) and solvent B (acetonitrile + 0.1% (v/v) formic acid). Peptides were scanned from 400 to 2000 *m/z* under positive ionization. The full scan was conducted at 70,000 resolution and data-dependent MS/MS scans at 175,000 resolution were performed on the resulting ions that exceeded a minimum threshold of 20,000. Peptide match was set as "preferred, exclude isotopes" and charge-state screening were enabled to reject singly and unassigned charged ions. Resulting peptide sequences were identified using MaxQuant 1.6.1.0 software (Supplemental Fig. S32).

In vitro ACP phosphopantetheinylation assay.
For standard phosphopantetheinylation reactions between non-actinobacterial *apo*-ACPs and Sfp R4-4, 500 μL reactions were performed in 50 mM phosphate buffer, pH 7.6 with concentrations of 200 μM ACP, 2.5 μM Sfp, 2.5 mM dithiothreitol (DTT), 2.5 mM CoA, and 10mM $MgCl_2$. Phosphopantetheinylation reactions between *apo*-gloACP and gloPPT were performed under similar conditions, with gloPPT in place of Sfp. gloPPT concentrations were varied between ~2.5 μM and ~10 μM in separate reactions. GloPPT phosphopantetheinylation reactions involving the addition of gloSCL and co-factors contained additional amounts of gloSCL, ATP, and sodium salicylate at concentrations of 3 μM, 5 mM, and 0.5 mM respectively.

Malonylation of *holo*-gloACP Q31G/T35L.
Malonylation protocols were adapted from Beltran-Alvarez *et al.*[60] All reaction components and buffers were kept in 50 mM phosphate buffer, pH 7.6. In brief, to a glass vial *holo*-gloACP Q31G/T35L and tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl) were added and pre-incubated at 30 °C for 30 minutes. Afterwards, malonyl-CoA, was added to reach final concentrations of 75 μM *holo*-ACP, 1 mM TCEP-HCl, and 1 mM malonyl-CoA in a total volume of 100 μL phosphate buffer. For reactions involving EcFabD/ScFabD, these MCATs were added to reach final concentrations of 1.5 μM. Reactions were incubated at 30 °C for 30 minutes, and malonyl-ACP was confirmed by LC-MS.

Starter unit priming of *holo*-gloACP Q31G/T35L using gloSCL.
Similar to malonylation reactions, all reaction components were kept in 50 mM phosphate buffer, pH 7.6. Reactions were initiated by pre-incubating *holo*-gloACP Q31G/T35L with TCEP-HCl for 30 minutes at 30 °C. Afterwards, gloSCL, sodium salicylate, $MgCl_2$, and ATP were added to reach final concentrations of 75 μM *holo*-ACP, 1 mM TCEP-HCl, 3 μM gloSCL, 0.5 mM sodium salicylate, 5 mM $MgCl_2$, and 5 mM ATP in a total volume of 200 μL phosphate buffer. Reactions were incubated at 30 °C for 2 hours and conversion to salicyl-gloACP Q31G/T35L was confirmed by LC-MS.

In vitro gloPKS product formation with native starter unit incorporation.
In vitro product formation protocols were adapted from Cheng *et al.*[44] All reaction components were kept 50 mM phosphate buffer, pH 7.6. In brief, *holo*-gloACP Q31G/T35L and TCEP-HCl

were first incubated for 30 minutes at 30 °C. Afterwards, gloSCL, sodium salicylate, $MgCl_2$, and ATP were added, and the reaction was incubated for an additional 30 minutes. Finally, ScFabD gloKS-CLF, glycerol, and malonyl-CoA/$^{13}C_3$-malonyl-CoA were added and the reaction was incubated for 3 hours at 30 °C. Final reaction concentrations were as follows: 75 μM *holo*-ACP, 1 mM TCEP-HCl, 1.5 μM gloSCL, 0.5 mM sodium salicylate, 5 mM $MgCl_2$, 5 mM ATP, 15 μM KS-CLF, 1.5 μM FabD, 10% (v/v) glycerol, and 2 mM malonyl-CoA/$^{13}C_3$-malonyl-CoA in a total volume of 300 μL. After 30 minutes, reactions were extracted twice with 1 mL of 99:1 EtOAc:AcOH. The extracts were combined, dried, resuspended in 300 μL MeOH, and analyzed by UPLC-MS following protocols described below.

***In vitro* gloPKS product formation with acetate priming.**
All reaction components were kept 50 mM phosphate buffer, pH 7.6. Reactions were initiated by incubating *holo*-gloACP Q31G/T35L and TCEP-HCl for 30 minutes at 30 °C. Afterwards, gloKS-CLF, ScFabD, glycerol, and malonyl-CoA/$^{13}C_3$-malonyl-CoA were added and the reaction was incubated for 3 hours at 30 °C. Final reaction concentrations were as follows: 75 μM *holo*-ACP, 1 mM TCEP-HCl, 15 μM KS-CLF, 1.5 μM FabD, 10% (v/v) glycerol, and 2 mM malonyl-CoA/$^{13}C_3$-malonyl-CoA in a total volume of 300 μL. After extraction, product formation was analyzed following identical protocols as listed in the previous product formation section.

Metabolite LC-MS characterization.
Molecules were analyzed on a Waters Acquity I-Class UPLC system coupled to a Synapt G2Si HDMS mass spectrometer in positive ion mode with a heated electrospray ionization (ESI) source in a Z-spray configuration. For small molecules, LC separation was performed on a Waters Acquity UPLC BEH C18 1.7 μm 2.1 x 50 mm column using an 0.6 mL/min gradient of 95/5 to 15/85 A/B over the course of four minutes. Eluent A is 0.1% (v/v) formic acid in water and B is 0.1% (v/v) formic acid in acetonitrile. Conditions on the mass spectrometer were as follows: capillary voltage 0.5 kV, sampling cone 40 V, source offset 80 V, source 120 °C, desolvation 250 °C, cone gas 0 L/h, desolvation gas 1000 L/h and nebulizer 6.5 bar. The analyzer was operated in resolution mode and low energy data was collected between 100 and 2000 Da at 0.2 sec scan time. Data was collected using a 20–40 V ramp trap collision energy. Masses were extracted from the TOF MS TICs using a 0.01 or 0.05 Da abs width.

Supplementary Information
The following Supporting Information is available free of charge at the ACS website.

Detailed Methods
Table S1. Homology comparison of thermorubin-producing proteins to *Gloeocapsa* homologs.
Table S2. Plasmid, primer, and amino acid sequences of all proteins used in this study.
Table S3. Thermocycler conditions for molecular cloning
Table S4. List of experimental molecular weights of the ACPs obtained by deconvoluting LC-MS results
Figure S1. SDS-PAGE of core *Gloeocapsa* proteins.
Figure S2. LC-MS spectrum of purified *apo*-gloACP (WT).
Figure S3. LC-MS spectrum of gloACP following coexpression with gloPPT.
Figure S4. LC-MS spectrum of *apo*-gloACP reacted *in vitro* with separately purified gloPPT.
Figure S5. LC-MS spectrum of coexpressed gloACP and gloPPT reacted *in vitro*.
Figure S6. SDS-PAGE and Western Blotting of gloPPT.

Figure S7. LC-MS spectrum of coexpressed gloACP and gloPPT reacted *in vitro* with gloSCL.
Figure S8. LC-MS spectrum of gloACP reacted with Sfp and Sfp R4-4.
Figure S9. Backbone NMR assignment and overlay of wild type gloACP and gloACP Q31G/T37L.
Figure S10. Chemical shift derived structural model of gloACP via CS-Rosetta
Figure S11. Full length multiple sequence alignment of various CPs and non-actinobacterial ACPs with varying Sfp-incompatibility.
Figure S12. LC-MS spectrum of purified *apo*-gloACP (Q31G/T35L).
Figure S13. LC-MS spectrum of *apo*-gloACP (Q31G/T35L) reacted *in vitro* with Sfp R4-4.
Figure S14. LC-MS spectrum of purified *apo*-gloACP (Q31G/T35I) reacted *in vitro* with Sfp R4-4.
Figure S15. LC-MS spectrum of purified *apo*-gloACP (Q31G) reacted *in vitro* with Sfp R4-4.
Figure S16. LC-MS spectrum of purified *apo*-gloACP (T35L) reacted *in vitro* with Sfp R4-4.
Figure S17. LC-MS spectrum of purified *apo*-dacACP.
Figure S18. LC-MS spectrum of dacACP reacted with Sfp and Sfp R4-4.
Figure S19. LC-MS spectrum of purified *apo*-panACP.
Figure S20. LC-MS spectrum of panACP reacted with Sfp and Sfp R4-4.
Figure S21. LC-MS spectrum of purified *apo*-dacACP (T43L).
Figure S22. LC-MS spectrum of dacACP (T43L) reacted with Sfp and Sfp R4-4.
Figure S23. LC-MS spectrum of purified *apo*-panACP (A30G/T34L/A37V).
Figure S24. LC-MS spectrum of panACP (A30G/T34L/A37V) reacted with Sfp and Sfp R4-4.
Figure S25. LC-MS spectrum of salicyl-gloACP (Q31G/T35L).
Figure S26. LC-MS spectrum of self-malonylated *holo*-gloACP (Q31G/T35L).
Figure S27. LC-MS spectrum of ScFabD catalyzed *holo*-gloACP (Q31G/T35L) malonlylation.
Figure S28. LC-MS spectrum of EcFabD catalyzed *holo*-gloACP (Q31G/T35L) malonylation.
Figure S29. Full extracted high resolution ion chromatograms of polyketide peaks of *m/z* = 529 and 547.
Figure S30. MS/MS fragmentation of single possible polyketide corresponding to previously hypothesized thermorubin precursor of *m/z* = 529.
Figure S31. Full extracted high resolution ion chromatograms of polyketide peaks of *m/z* = 451, 469, 475, 493, and 511.
Figure S32. Tandem proteolysis analysis of gloKS-CLF.

References

1. Newman, D. J. & Gordon, C. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).

2. Stallforth, P., Mittag, M., Brakhage, A. A., Hertweck, C. & Hellmich, U. A. Functional modulation of chemical mediators in microbial communities. *Trends Biochem. Sci.* **48**, 71–81 (2023).

3. Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat Prod Rep* **24**, 162–190 (2007).
4. Genilloud, O. & Vicente, F. Tetracycline Antibiotics and Novel Analogs. in *Antimicrobials: New and Old Molecules in the Fight Against Multi-resistant Bacteria* (eds. Marinelli, F. & Genilloud, O.) 231–245 (Springer Berlin Heidelberg, 2014). doi:10.1007/978-3-642-39968-8_12.
5. Sulpizio, A., Crawford, C. E. W., Koweek, R. S. & Charkoudian, L. K. Probing the structure and function of acyl carrier proteins to unlock the strategic redesign of type II polyketide biosynthetic pathways. *J. Biol. Chem.* **296**, 100328 (2021).
6. Tsai, S.-C. (Sheryl). The Structural Enzymology of Iterative Aromatic Polyketide Synthases: A Critical Comparison with Fatty Acid Synthases. *Annu. Rev. Biochem.* **87**, 503–531 (2018).
7. Sundaram, S. & Hertweck, C. On-line enzymatic tailoring of polyketides and peptides in thiotemplate systems. *Curr. Opin. Chem. Biol.* **31**, 82–94 (2016).
8. Roessner, C. A. & Scott, A. I. Achieving natural product synthesis and diversity via catalytic networking ex vivo. *Chem. Biol.* **3**, 325–330 (1996).
9. Shen, B. & Hutchinson, C. R. Enzymatic Synthesis of a Bacterial Polyketide from Acetyl and Malonyl Coenzyme A. *Science* **262**, 1535–1540 (1993).
10. McDaniel, R., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. *Nature* **375**, 549–554 (1995).
11. McDaniel, R., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. Engineered Biosynthesis of Novel Polyketides. *Science* **262**, 1546–1550 (1993).

12. Fu, H., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. Engineered Biosynthesis of Novel Polyketides: Dissection of the Catalytic Specificity of the act Ketoreductase. *J. Am. Chem. Soc.* **116**, 4166–4170 (1994).
13. Fu, H., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. Relaxed Specificity of the Oxytetracycline Polyketide Synthase for an Acetate Primer in the Absence of a Malonamyl Primer. *J. Am. Chem. Soc.* **116**, 6443–6444 (1994).
14. Hopwood, D. A. *et al.* Production of 'hybrid' antibiotics by genetic engineering. *Nature* **314**, 642–644 (1985).
15. Shen, B., Summers, R. G., Wendt-Pienkowski, E. & Hutchinson, C. R. The Streptomyces glaucescens tcmKL polyketide synthase and tcmN polyketide cyclase genes govern the size and shape of aromatic polyketides. *J. Am. Chem. Soc.* **117**, 6811–6821 (1995).
16. Fu, H. & Khosla, C. Antibiotic activity of polyketide products derived from combinatorial biosynthesis: Implications for directed evolution. *Mol. Divers.* **1**, 121–124 (1996).
17. Zawada, R. J. & Khosla, C. Heterologous expression, purification, reconstitution and kinetic analysis of an extended type II polyketide synthase. *Chem. Biol.* **6**, 607–615 (1999).
18. Gao, X., Wang, P. & Tang, Y. Engineered polyketide biosynthesis and biocatalysis in Escherichia coli. *Appl. Microbiol. Biotechnol.* **88**, 1233–1242 (2010).
19. Hillenmeyer, M. E., Vandova, G. A., Berlew, E. E. & Charkoudian, L. K. Evolution of chemical diversity by coordinated gene swaps in type II polyketide gene clusters. *Proc. Natl. Acad. Sci.* **112**, 13952 (2015).

20. Chen, S., Zhang, C. & Zhang, L. Investigation of the Molecular Landscape of Bacterial Aromatic Polyketides by Global Analysis of Type II Polyketide Synthases. *Angew. Chem. Int. Ed.* **61**, e202202286 (2022).

21. McBride, C. M., Miller, E. L. & Charkoudian, L. K. An updated catalogue of diverse type II polyketide synthase biosynthetic gene clusters captured from large-scale nucleotide databases. *Microbial Genomics* vol. 9 (2023).

22. Cummings, M. *et al.* Assembling a plug-and-play production line for combinatorial biosynthesis of aromatic polyketides in Escherichia coli. *PLOS Biol.* **17**, 1–34 (2019).

23. Klein, J. G., Wu, Y., Kokona, B. & Charkoudian, L. K. Widening the bottleneck: Heterologous expression, purification, and characterization of the Ktedonobacter racemifer minimal type II polyketide synthase in Escherichia coli. *Bioorg. Med. Chem.* **28**, 115686 (2020).

24. Sugimoto, Y. *et al.* A metagenomic strategy for harnessing the chemical repertoire of the human microbiome. *Science* **366**, eaax9176 (2019).

25. Lambalot, R. H. & Walsh, C. T. Cloning, Overproduction, and Characterization of the Escherichia coli Holo-acyl Carrier Protein Synthase (∗). *J. Biol. Chem.* **270**, 24658–24661 (1995).

26. Sunbul, M., Marshall, N. J., Zou, Y., Zhang, K. & Yin, J. Catalytic Turnover-Based Phage Selection for Engineering the Substrate Specificity of Sfp Phosphopantetheinyl Transferase. *J. Mol. Biol.* **387**, 883–898 (2009).

27. Quadri, L. E. N. *et al.* Characterization of Sfp, a Bacillus subtilis Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases. *Biochemistry* **37**, 1585–1595 (1998).

28. Liu, X. *et al.* Heterologous Biosynthesis of Type II Polyketide Products Using E. coli. *ACS Chem. Biol.* **15**, 1177–1183 (2020).

29. Yang, D., Jang, W. D. & Lee, S. Y. Production of Carminic Acid by Metabolically Engineered Escherichia coli. *J. Am. Chem. Soc.* **143**, 5364–5377 (2021).

30. Bräuer, A. *et al.* Structural snapshots of the minimal PKS system responsible for octaketide biosynthesis. *Nat. Chem.* **12**, 755–763 (2020).

31. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of *E. coli*. *Science* **291**, 1790–1792 (2001).

32. Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci.* **105**, 4685–4690 (2008).

33. Tufar, P. *et al.* Crystal Structure of a PCP/Sfp Complex Reveals the Structural Basis for Carrier Protein Posttranslational Modification. *Chem. Biol.* **21**, 552–562 (2014).

34. Aragozzini, F., Craveri, R., Maconi, E., Ricca, G. S. & Scolastico, C. Thermorubin biosynthesis: evidence for the involvement of both salicylic acid and an undecaketide. *J Chem Soc Perkin Trans 1* 1865–1867 (1988) doi:10.1039/P19880001865.

35. McCord, J. P., Kohanov, Z. A. & Lowell, A. N. Thermorubin Biosynthesis Initiated by a Salicylate Synthase Suggests an Unusual Conversion of Phenols to Pyrones. *ACS Chem. Biol.* **17**, 3169–3177 (2022).

36. Florova, G., Kazanina, G. & Reynolds, K. A. Enzymes Involved in Fatty Acid and Polyketide Biosynthesis in Streptomyces glaucescens: Role of FabH and FabD and Their Acyl Carrier Protein Specificity. *Biochemistry* **41**, 10462–10471 (2002).

37. Serre, L., Verbree, E. C., Dauter, Z., Stuitje, A. R. & Derewenda, Z. S. The Escherichia coli Malonyl-CoA:Acyl Carrier Protein Transacylase at 1.5-Å Resolution.: CRYSTAL STRUCTURE OF A FATTY ACID SYNTHASE COMPONENT ∗. *J. Biol. Chem.* **270**, 12961–12964 (1995).

38. Arthur, C. J. *et al.* Self-Malonylation Is an Intrinsic Property of a Chemically Synthesized Type II Polyketide Synthase Acyl Carrier Protein,. *Biochemistry* **44**, 15414–15421 (2005).

39. Hertweck, C. *et al.* Context-Dependent Behavior of the Enterocin Iterative Polyketide Synthase: A New Model for Ketoreduction. *Chem. Biol.* **11**, 461–468 (2004).

40. Jackson, D. R. *et al.* Structural Insights into Anthranilate Priming during Type II Polyketide Biosynthesis. *ACS Chem. Biol.* **11**, 95–103 (2016).

41. Moore, B. S. & Hertweck, C. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat Prod Rep* **19**, 70–99 (2002).

42. Tang, Y., Lee, T. S. & Khosla, C. Engineered Biosynthesis of Regioselectively Modified Aromatic Polyketides Using Bimodular Polyketide Synthases. *PLOS Biol.* **2**, null (2004).

43. Xu, Z., Schenk, A. & Hertweck, C. Molecular Analysis of the Benastatin Biosynthetic Pathway and Genetic Engineering of Altered Fatty Acid−Polyketide Hybrids. *J. Am. Chem. Soc.* **129**, 6022–6030 (2007).

44. Cheng, Q., Xiang, L., Izumikawa, M., Meluzzi, D. & Moore, B. S. Enzymatic total synthesis of enterocin polyketides. *Nat. Chem. Biol.* **3**, 557–558 (2007).

45. Keatinge-Clay, A. T., Maltby, D. A., Medzihradszky, K. F., Khosla, C. & Stroud, R. M. An antibiotic factory caught in action. *Nat. Struct. Mol. Biol.* **11**, 888–893 (2004).

46. Szu, P.-H. *et al.* Analysis of the Ketosynthase-Chain Length Factor Heterodimer from the Fredericamycin Polyketide Synthase. *Chem. Biol.* **18**, 1021–1031 (2011).

47. Leblanc, C. *et al.* 4′-Phosphopantetheinyl Transferase PptT, a New Drug Target Required for Mycobacterium tuberculosis Growth and Persistence In Vivo. *PLOS Pathog.* **8**, 1–16 (2012).

48. Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P. & Burkart, M. D. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat Prod Rep* **31**, 61–108 (2014).

49. Tang, Y., Tsai, S.-C. & Khosla, C. Polyketide Chain Length Control by Chain Length Factor. *J. Am. Chem. Soc.* **125**, 12708–12709 (2003).

50. Saha, S. K., Uma, L. & Subramanian, G. An Improved Method for Marine Cyanobacterial DNA Isolation. *World J. Microbiol. Biotechnol.* **21**, 877–881 (2005).

51. Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, W81–W87 (2019).

52. Gibson, D. G. *et al.* Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **329**, 52–56 (2010).

53. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).

54. Azatian, S. B., Kaur, N. & Latham, M. P. Increasing the buffering capacity of minimal media leads to higher protein yield. *J. Biomol. NMR* **73**, 11–17 (2019).

55. Wishart, D. S. *et al.* 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **6**, 135–140 (1995).

56. Skinner, S. P. *et al.* CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR* **66**, 111–124 (2016).

57. Maciejewski, M. W. *et al.* NMRbox: A Resource for Biomolecular NMR Computation. *Biophys. J.* **112**, 1529–1534 (2017).

58. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013).

59. Winkler, R. ESIprot: a universal tool for charge state determination and molecular weight calculation of proteins from electrospray ionization mass spectrometry data. *Rapid Commun. Mass Spectrom.* **24**, 285–294 (2010).

60. Beltran-Alvarez, P., Cox, R. J., Crosby, J. & Simpson, T. J. Dissecting the Component Reactions Catalyzed by the Actinorhodin Minimal Polyketide Synthase. *Biochemistry* **46**, 14672–14681 (2007).

Acknowledgements/Funding

We are grateful for generous support from the National Science Foundation (CAREER CHE-1652424 and CHE-2201984) and the Henry Dreyfus Teacher Scholar Award (TH-10-2020) to LKC. We thank the Wistar Institute for conducting tandem proteolysis experiments and Professor Dominic Campopiano (University of Edinburgh) for providing the FabD expression plasmid. This work was also supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the collaborative research center SFB1127/3 ChemBioSys, project ID 3974852 and the Cluster of Excellence "Balance of the Microverse" EXC 2051 – Project-ID 390713860 (to UAH). UAH acknowledges an instrumentation grant for a high-field NMR spectrometer by the REACT-EU EFRE Thuringia (Recovery assistance for cohesion and the territories of Europe, European Fonds for Regional Development, Thuringia) initiative of the European Union. We are thankful to the Research Corporation for Scientific Advancement (Cottrell Scholars) for funding and supporting the collaboration between LKC and UAH.

Author information

Authors and Affiliations.

Department of Chemistry, Haverford College, Haverford, PA, U.S.A.
Yae In Cho, Kevin Li, Rebecca S. Koweek, Grayson Hamrick, Ngọc Khánh Hoàng, Margaret A. Bowen, Bashkim Kokona, and Louise K. Charkoudian

Faculty of Chemistry and Earth Sciences, Institute for Organic Chemistry and Macromolecular Chemistry, Friedrich Schiller University Jena, Jena, Germany
Ute A. Hellmich, Mai Anh Tran and Christoph Wiedemann

Cluster of Excellence Balance of the Microverse, Friedrich Schiller University Jena, Jena, Germany
Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University, Frankfurt, Germany
Ute A. Hellmich

Department of Microbiology & Immunology, Center for Advanced Microbial Processing, Institute for Molecular Medicine and Infectious Disease. Drexel University College of Medicine, Philadelphia, PA, U.S.A.
Joris Beld

Corresponding authors.
Ute A. Hellmich and Louise K. Charkoudian