Mutagenesis supports AlphaFold prediction of how modular polyketide synthase acyl carrier proteins dock with downstream ketosynthases

Adrian Keatinge-Clay[1], Melissa Hirsch[2], Ronak R. Desai[1], and Shreyas Annaswamy[1]

[1]The University of Texas at Austin Department of Molecular Biosciences
[2]The University of Texas at Austin Department of Chemistry

March 31, 2024

Abstract

The docking of an acyl carrier protein (ACP) domain with a downstream ketosynthase (KS) domain in each module of a polyketide synthase (PKS) helps ensure accurate biosynthesis. If the polyketide chain bound to the ACP has been properly modified by upstream processing enzymes and is compatible with gatekeeping residues in the KS tunnel, a transacylation reaction can transfer it from the 18.1-Å phosphopanthetheinyl arm of the ACP to the reactive cysteine of the KS. AlphaFold-Multimer predicts a general interface for these transacylation checkpoints. Half of the solutions obtained for 50 ACP/KS pairs show the KS motif TxLGDP forming the first turn of an a-helix, as in reported structures, while half show it forming a type I b-turn not previously observed. Solutions with the latter conformation may represent how these domains are relatively positioned during the transacylation reaction, as the entrance to the KS active site is relatively open and the phosphopantetheinylated ACP serine and the reactive KS cysteine are relatively closer - 17.2 Å vs. 20.9 Å, on average. To probe the predicted interface, 20 mutations were made to KS surface residues within the model triketide lactone synthase **P1- P6- P7**. The activities of these mutants are consistent with the proposed interface.

INTRODUCTION

Modular polyketide synthases (PKS's) are multidomain, enzymatic assembly lines that play the major role in the production of important pharmaceuticals, such as the antibiotic erythromycin and the anticancer agent epothilone.[1-3] Each module is minimally comprised of an acyl carrier protein (ACP) domain that shuttles polyketide intermediates between enzymatic domains using a phosphopantetheinyl arm and a ketosynthase (KS) domain that selects which intermediates to extend and pass along the assembly line (Figure 1). A module may also contain an acyltransferase (AT) domain that selects α-carboxyacyl extender units, such as a malonyl or methylmalonyl group, and processing domains, such as a ketoreductase (KR), dehydratase (DH), or enoylreductase (ER), that perform chemistry on extended intermediates. While the high-resolution structures of each of these domains have been known for many years, the interfaces between them are only now being characterized.[4]

Each KS of an assembly line is thought to possess distinct docking sites for the upstream and downstream ACP's with which it collaborates.[5, 6] It was hypothesized that the upstream ACP docks at a transacylation site to present intermediates for transfer to the reactive KS cysteine and the downstream ACP docks at an extension site to collect the intermediate through a decarboxylative Claisen-like condensation with a bound α-carboxyacyl extender unit. The extension site has been revealed by cryogenic-electron microscopy (cryo-EM) and x-ray crystallography.[7-9] In cryo-EM studies of both PikAIII (the third polypeptide of the Pikromycin synthase)[10] and *Mycobacterium smegmatis*Pks13 (PDB: 8CUY, 8CV1),[11] other ACP/KS interactions were observed; however, in each of these interactions the phosphopantetheinylated serine of ACP is more than 28 Å

from the KS reactive cysteine. As the thioester carbon attacked by the KS reactive cysteine can only stretch 19.2 Å from the phosphopantetheinylated serine oxygen, none of these interfaces represent the transacylation site.

Improvements in folding/docking algorithms are enabling the reliable prediction of physiologically relevant domain-domain interfaces.[12-15] Since ACP's and downstream KS's collaborate within the PKS module (updated boundary used[16, 17]) to gatekeep for processed polyketide intermediates,[18] evolutionarily co-migrate,[16, 17, 19, 20] and are best maintained together within engineered assembly lines,[21-24] we hypothesized that modern folding/docking algorithms would detect the interface between these domains.

The AlphaFold-Multimer[14] predictions for 50 natural ACP/KS pairs from diverse, well-characterized PKS's are highly similar.[2] Half of the top solutions contain a conformational change of the KS TxLGDP motif at the putative docking interface that enables a closer approach (17.2 Å vs. 20.9 Å, on average) of the phosphopantetheinylated ACP serine and the reactive KS cysteine. These solutions provide a molecular description of how acyl-ACP's dock with KS's to enable the transacylation reaction. A model triketide lactone synthase composed of the 1st, 6th, and 7th modules of the Pikromycin synthase, **P1 -P6 -P7** , was employed to test the predicted transacylation site.[24, 25] The relative decreases in the activities of 20 variants containing mutations to surface residues on the KS from the 6thmodule, PikKS6, are consistent with the proposed interface. The data suggest that two residues conserved on the surface of KS, an asparagine in position 275 and a leucine in position 315, play major roles in binding acyl-ACP's for transacylation.

RESULTS

AlphaFold-Multimer predictions for 50 ACP+KS dimers

In the majority of *cis* -AT PKS modules, the ACP and KS domains are embedded in the same polypeptide; however, modules are also frequently split such that ACP is embedded in the upstream polypeptide and KS in the downstream polypeptide[20] . C- and N-terminal docking domain motifs (CDD & NDD) at the polypeptide termini help noncovalently connect the domains from these modules[26, 27] . We selected 45 of 50 of the ACP+KS's in this study from modules naturally embedded in a single polypeptide, since the linker connecting these domains contains fewer than 15 residues and provides a restraint for the folding/docking of these domains[20] . Five pairs of ACP's and KS's naturally embedded in separate polypeptides were also investigated by connecting them with a 15-residue polyglycine linker (G_{15}). As KS's are dimeric, 2 copies of each of the 50 ACP+KS sequences were provided to AlphaFold-Multimer (Data File 1, see Table S1 for PKS abbreviations, Figures S1-S3). For each ACP+KS, 5 energy-minimized solutions were obtained (Data Files 2-51)[14] .

The ACP's consistently dock to the same location, with the serine that becomes phosphopantetheinylated facing the entrance to the KS active site (Figure 2a and S2). ACP and KS bind *in trans* in 76% of the solutions, with the ACP domains almost exclusively making contacts with residues from the KS of the opposite monomer. As the N-terminal end of αI' of ACP (secondary-structure nomenclature,[20, 28] prime used to distinguish ACP from KS, numbering from PikACP6 and PikKS6 used to discuss all ACP/KS pairs) is predicted to be close to the AT domain downstream of KS, the structures of 5 ACP+KS+AT dimers [Ery(ACP1+KS1+AT2), Ery(ACP4+KS4+AT5), Pik(ACP1+KS1+AT2), Pik(ACP6+KS6+AT7), and Rap(ACP1+KS1+AT2)] were also predicted (Figures 2b and S4, Data Files 52-57). All of the solutions show ACP bound to the same site as in the solutions for the ACP+KS dimers with no contact between ACP and AT. Either a rotation about αIII' or a hinge motion between KS and AT would be necessary for ACP and the downstream AT to make contact.[7, 8]

The interface between ACP and KS is similar in all solutions (Figures 3 and S2). While contacts with ACP residues in and around αIII' (residues 73'-80') are most substantial, contacts with residues in a one-turn helix upstream of αII' (residues 46'-48') and residues at the N-terminal end of αI' (residues 53'-60') are also significant. KS residues involved in the contacts include those in the structured loop between β8 and α10 (residues 273-283), the N-terminal end of α11 (residues 314-322), and the loop between β11 and β12 (residues 381-388).

Solutions shows a KS motif in two conformations at the interface

In half of the 50 top solutions, the KS TxLGDP motif (residues 313-318) adopts a type I β-turn distinct from the α-helical conformation observed in solved structures and the other solutions (Figure 4).[7, 8, 29-32] In 17 of these 25 solutions (Data Files 6, 8, 11, 12, 16, 18, 22, 24, 25, 27, 36, 37, 43, 44, 45, 46, 50), the β-turn conformation is associated with a relative rotation of ACP about αIII' such that L315 of the TxLGDP motif makes contact with the serine that becomes phosphopantetheinylated, S53' (L315 C_γ - S53' $C_\beta < 5.0$ Å). To investigate whether formation of the β-turn is associated with ACP binding, the sequences of the KS's from the 25 top solutions with the β-conformation were provided to AlphaFold-Multimer. None of the predictions from the predicted KS dimers contain the TxLGDP motif in the β-conformation.

For 21 of the ACP+KS's, the 5 predictions made by AlphaFold include both a solution in which the TxLGDP motif is in the α-conformation and a solution in which the motif is in the β-conformation and ACP is relatively rotated (Data Files 2, 6, 11, 12, 13, 14, 16, 18, 22, 27, 31, 32, 33, 36, 39, 43, 44, 45, 48, 50, 51, 58-78). Thus, the differences between the conformations of this motif as well as the relative orientations of ACP and KS can be directly compared (Figures 3-4). In the β-conformation of the TxLGDP motif of PikKS6, a hydrogen bond is present between the carbonyl of T313 and the NH of G316, rather than the NH of D317, as in the α-conformation. The D317 carboxylate associates with the NH's of T313 and an invariant glycine (G312) immediately upstream of the motif as well as the T313 hydroxyl. The oxygen of the T313 hydroxyl and the D317 NH form another hydrogen bond. P318 adopts the γ-endo conformation rather than the γ-exo conformation in the α-conformation. The β-conformation is stabilized through van der Waals interactions with a structured loop (A232-G234) as well as a hydrogen bond between the L315 carbonyl and the hydroxyl group of a serine or threonine at position 381. The conformational shift is largest for L315 and G316, their C_α's shifting 4.7 Å and 7.0 Å, respectively.

The conformational change could be important in allowing the approach of the KS reactive cysteine and the thioester of the acylated phosphopantetheinyl arm. In the 17 top solutions containing the β-conformation and the rotated ACP, the arm can move closer to the KS reactive cysteine. While the average distance between the ACP serine oxygen and the KS cysteine sulfur is 20.9 Å for the solutions containing the α-conformation (excluding AjuMod3, *vide infra*), it is 17.2 Å (3.7 Å closer, on average) for these 17 solutions (Table S2). Since the distance between the side chain oxygen of the phosphopantetheinylated serine and the carbon of the thioester carbonyl of an acyl-ACP is maximally 19.2 Å, the β-conformation may be more relevant to the transacylation reaction. An acetyl-phosphopantetheinyl moiety modeled with the program Coot[33] into a PikMod6 solution with the β-conformation shows that its *gem* -dimethyl moiety can hydrophobically interact with P280, that its distal carbonyl can coordinate with the H309 $N_\varepsilon H$ and the T311 hydroxyl group, and that its thioester carbonyl can insert into the oxyanion hole formed by the NH's of C174 and V217 (Figure 3c) (Data File 79). The conformational change of the TxLGDP motif provides space for the phosphopantetheinyl arm and allows the approach of the thioester and the reactive cysteine to enable the transacylation reaction.

Complementarity at the transacylation site

The interactions observed in the top solution for PikMod6 are representative of interactions observed in each of the 17 top solutions with the rotated ACP (L315 contacting αII') (Table S2). As solutions with an unrotated ACP (24 with the α-conformation and 8 with the β-conformation) generally do not permit the thioester connecting the phosphopantetheinyl arm and the polyketide to approach the KS reactive cysteine, they are not considered here. All contacts are made between residues in a 36-aa region of PikACP6 (positions 45'-80') and in a 109-aa region of PikKS6 (positions 275-383) (Figure 3a).

The solution from PikMod6 shows the T74' methyl group from its ACP occupying a hydrophobic pocket formed by P318, I319, and Q322 on the surface of PikKS6 (Figures 3c and 5a-b, Table S2). A threonine plays this role in 15 out of the 17 top solutions with the rotated ACP. The T74' hydroxyl group and the D78' carboxylate are observed making a hydrogen bond in each solution. The sidechains of residues at positions 72' and 73' also make contact with KS. A proline most commonly at position 72' usually makes Van der Waals contact with the sidechain of the residue in position 322, and an alanine most commonly at position

73' fills space adjacent to the backbone carbonyl of position 383. When a serine or threonine is in position 73', its side chain usually shares a hydrogen bond with backbone carbonyl of the residue in position 383 (EbeMod7, CpkMod4, E837Mod2, EryMod3). An unusual leucine in position 73' of BafMod3 makes greater van der Waals contact with the residue in position 383.

In the PikMod6 solution, D78' interacts with the N-terminal end of α10. A hydrogen bond is present between its carboxylate and the NH of G282, while the nonpolar portion of D78' makes van der Waals interactions with P283. A hydrogen bond between the carboxylate of the aspartate in position 78' with the NH of a glycine at position 282 is present in the 17 top solutions with the rotated ACP. The backbone carbonyl of D78' most commonly forms a hydrogen bond with the side chain of an asparagine in position 281.

In 16 of the 17 top solutions with the rotated ACP, a phenylalanine occupies position 77', while in EryMod1 a tyrosine is present. In all of the solutions, the benzyl group is equivalently oriented, with its methylene in contact with the δ-methyl group of the isoleucine at position 319. When the phosphopantetheinyl arm is present, its*gem* -dimethyl moiety can hydrophobically interact with the aromatic ring. In each of the 25 top solutions with the α-conformation, it is the *gem* -dimethyl moiety of L315 that interacts with the aromatic ring.

The C-terminal end of αIII' and the N-terminal end of a one-turn helix (residues 46'-49') upstream of αII' in ACP make contact with the KS surface residue at position 275 and nearby residues in the structured loop between β8 and α10. Van der Waals contacts between a proline in position 80' and residues in positions 274 and 275 are substantial. This proline is much less conserved in the ACP's of thioesterase (TE)-containing modules in which KS is replaced by a TE (Figure S5). Usually, 1 or 2 hydrogen bonds are formed between the carbonyls of residues in positions 76' and 77' and the NH_2 of an asparagine highly conserved at position 275. In the absence of these hydrogen bonds, one is usually formed between the asparagine sidechain carbonyl and the backbone NH of the residue in position 47'. The residue in position 47' is usually an arginine or lysine, whose positively-charged group is most often observed in contact with the sidechain oxygen of a threonine conserved at position 278. In contrast, the residue in position 47' is usually hydrophobic in the ACP's of TE-containing modules. In PikMod3, the residue at position 275 is a serine rather than an asparagine. The serine sidechain makes shape-complementary interactions with ACP and its backbone carbonyl forms a hydrogen bond with the backbone NH of an unusual leucine in position 47'. The A279 side chain and the sidechain of this leucine also make hydrophobic contact.

Although αIII' of ACP makes the most contact with the KS surface, αII' is also involved in significant interactions. In the PikMod6 solution, the methylene of S53' and the sidechains of L54' and V57' on the ACP create a hydrophobic pocket for L315 from the KS TxLGDP motif. Residues in αII' also make contact with the residue in position 383, most often with its sidechain, but also with its backbone carbonyl. The residue in position 57', usually a valine, often makes contact with the sidechain of the residue in position 383, most commonly through van der Waals interactions with a histidine or glutamine. In 7 of the 17 top solutions with the rotated ACP the sidechain of a conserved arginine at position 70' in the loop between αII' and αIII' makes contact with residues in a 3_{10} helix on KS (positions 385-387), especially the aspartate conserved in position 385.

Testing the transacylation site through mutations to KS surface residues

To assess the proposed transacylation site, PikKS6 surface residues were mutated within **P1 -P6 -P7** , a triketide synthase composed of PikMod1, PikMod6, and PikMod7 (Figure 5b).[24, 25] In the top solution for the association of PikACP6 and PikKS6, the PikKS6 surface residues N275, H281, L315, I319, Q322, D382, and D385 make favorable contacts with PikACP6, while Q217, S230, Q286, S300, R314, Q330, E379, S387, E392, and T395 make little or no contact. Each of these residues was individually mutated to alanine within**P1 -P6 -P7** .[34] The prominent putative interface residues N275 and L315 were also individually mutated to aspartate and valine, respectively. As a control, a key residue at the condensation site, T140, was individually mutated to alanine.[7, 8]**P1 -P6 -P7** and its 20 point mutants were expressed in *Escherichia coli* K207-3,[35]and triketide lactone production was measured by LC/MS after 8 days. The observed decreases

in production are consistent with the predicted transacylation site.

DISCUSSION

Among the many docking events within a polyketide assembly line, those between KS's and their upstream ACP's are of high significance because they enable KS's to gatekeep for properly processed intermediates and thus ensure the fidelity of the synthase[18] . Although modern structural biology techniques, such as cryo-EM, have advanced our knowledge of the interactions between PKS domains, they are still resource-intensive and low throughput.[7, 8] In contrast, modern *in silico* methods require few resources and are high throughput. Thus, we sought to identify the transacylation site using the most advanced folding/docking algorithms.

Alphafold-Multimer uses restraints derived from covariation analysis, both within and between domains, to guide folding/docking.[15] In half of the 50 top solutions for the ACP+KS dimers, a structural feature is present at the interface that has not been previously observed - the TxLGDP motif of KS forms a type I β-turn rather than the α-helical turn present in reported structures and the other 25 top solutions. The β-conformation is correlated with a closer approach between the ACP and KS domains. The distance between the side-chain oxygen of the conserved ACP serine and the reactive KS cysteine is greater than 18.4 Å in 24 of the 25 top solutions with the α-conformation (20.8 Å, on average). The exception is AjuMod3 (15.8 Å), a module containing non-ribosomal peptide synthetase (NRPS) domains,[36] in which an unusual phenylalanine in position 47' makes contact with an unusual valine at position 275. The 17 top solutions that possess the β-conformation and a relatively rotated ACP possess distances between the phosphopantetheinylated serine and the reactive cysteine short enough for polyketides to transfer from the acyl-ACP to the KS (17.2 Å, on average).

The TxLGDP motif has not been experimentally observed in the β-conformation. However, these studies suggest that acyl chains are not readily transferred between ACP and KS when this motif is in the α-conformation. This raises the possibility that favorable interactions between the polyketide intermediate and the KS substrate tunnel facilitate the transition from the α-conformation to the β-conformation, which enables a closer association of the acyl-ACP and KS, productive interactions between KS and the phosphopantetheinyl arm (such as the coordination of H309 $N_{\varepsilon}H$ and T311 OH with the distal carbonyl), and polyketide transfer through the approach of the thioester and the reactive cysteine. This may be part of the KS gatekeeping mechanism by which properly processed intermediates are selected for transfer to the reactive cysteine.[18]

During the reaction cycle of a module, ACP associates with the upstream AT and processing domains as well as the downstream KS (Figure 1). While it must be complementary to each of these cognate enzymes, how well ACP's dock with noncognate domains introduced through natural recombination events or genetic engineering is unknown. As the domains of one module are able to evolutionarily co-migrate separately from the domains of other modules, their interfaces could diverge. If this is the case, engineering efforts should focus on module-swapping rather than domain-swapping. In a recent KS gatekeeping study, AmpKS15 was swapped for PikKS3 in PikMod3 of**P1** -**P2** -**P3** -**P7** , a tetraketide synthase composed of PikMod1, PikMod2, PikMod3, and PikMod7.[37] A triketide rather than the expected tetraketide was produced by the resulting synthase, indicating that the hybrid third module was completely skipped. An analysis of natural ACP/KS interfaces of AmpMod15 and PikMod3 reveals large differences in how their ACP's interact with the KS surface residue in position 275 (asparagine in AmpKS15 and serine in PikKS3).

The docking solutions presented here show interfaces between KS's and their upstream ACP's that are quite distinct from one another (Figure 5, Table S2). The most dramatic example is the interface between AjuACP3 and AjuKS3, in which a valine occupies position 275 of AjuKS3 and AjuACP3 is rotated 17º relative to the consensus orientation. As mentioned above, KS's containing a serine at position 275, such as PikKS3 and VstKS7, may not be recognized by most ACP's since they associate with KS's containing an asparagine at this position.

The transacylation site may be under selective pressure to diversify, whereas the extension site may be under selective pressure not to. The ordered assembly of PKS polypeptides could be jeopardized if the interfaces

between KS's and their upstream ACP's were too similar. CDD' and NDD's do not make the only contact between PKS polypeptides, since ACP's and KS's also associate, at least while mediating the transacylation reaction and likely even when ACP is unacylated.[27] The surface area for the transacylation site is larger than that observed for CDD/NDD interfaces (1338 Å^2 for PikACP6 with PikKS6 vs. 1043 Å^2 for one DEBS2 CDD helix with the dimeric DEBS3 NDD).[26] If all the KS's and ACP's at the polypeptide termini equally docked with one another, the synthase could assemble incorrectly. In contrast, the evolution of polyketide assembly lines may benefit from the interface between KS's and downstream ACP's not diverging, such that a module still functions when moved to a new location in the same or a different synthase.[7, 8] In the new location, the ACP and KS domains must respectively collaborate with a new upstream KS and downstream ACP to catalyze extension reactions.

The proposed transacylation site is consistent with the observed losses in activity from the 20 **P1 -P6 -P7** point mutants (Figure 5b). That mutations to the conserved KS surface residues N275 and L315 resulted in the least active synthases agrees with the interactions between PikACP6 and PikKS6 proposed to enable the transacylation reaction. The AlphaFold-Multimer predictions show the sidechain of N275 making substantial contact with PikACP6, through Van der Waals contact with P80' and hydrogen bond contact with both the C-terminal end of αIII' and the N-terminal end of the one-turn helix upstream of αII'. The AlphaFold-Multimer predictions also show the invariant leucine in the TxLGDP motif at position 315 making significant hydrophobic contact with F77' of PikACP6 when the motif is in the α-conformation and with a hydrophobic pocket on αII' next to the phophospantheinylated serine when the motif is in the β-conformation. Each of the other mutations also resulted in losses of activity compared to unmutated **P1 -P6 -P7** , usually moreso when located at or near the putative interface (*e.g.* , I319A, H281A, Q286A, D382A, D385A, Q330A). The relatively small decreases in activity observed for the R314A and Q322A point mutants might be surprising given the proximity of these residues to the interface; however, in the Alphafold-Multimer predictions the arginine and the glutamine sidechains make minimal Van der Waals contact with PikACP6. The low activity of the T140A point mutant control was expected from the importance of this residue at the PikKS6 extension site.[7, 8]

AlphaFold has apparently helped discover the transacylation site on the surface of KS and provided 50 examples of how upstream ACP's dock at that interface. Such a feat through experimental structural biology techniques would require many more resources and much more time. AlphaFold will likely help predict other domain-domain interactions within PKS's, although these should be experimentally tested within functional assembly lines. Our burgeoning understanding of how ACP domains dock with cognate enzymes is illuminating the dynamics and architectures of PKS's and will help realize the potential of these assembly lines in producing designer molecules and medicines.

METHODS

AlphaFold-Multimer predictions

AlphaFold-Multimer (version 3) was run on Google Colab servers. All sequences begin 55 residues upstream of the phosphopantetheinylated serine of ACP. ACP+KS sequences end 10 residues after the KS GTNAH motif, at the updated module boundary[17, 20] , and ACP+KS+AT sequences end 3 residues after the LPTYxFxxxxYWL motif.

Constructing and culturing P1-P6-P7 point mutants

The second of the 2 plasmids encoding the**P1 -P6 -P7** triketide lactone synthase, pTM5, was used as a template to generate 20**P1 -P6 -P7** point mutants.[24] Two halves of the plasmid were amplified using 2 pairs of mutagenic primers and joined through Gibson assembly (Table S3). The plasmids encoding**P1 -P6 -P7** and each of its point mutants were co-transformed into *E. coli* K207-3.[35] A single colony was used to inoculate 5 mL of Luria–Bertani (LB) media containing 100 mg L^{-1} kanamycin and 100 mg L^{-1} streptomycin. After the starter culture grew for 16 h at 37 °C and 250 rpm, 400 μL of it was used to innoculate 250-mL Erlenmeyer flasks containing 40 mL production media [5 g L^{-1} yeast extract, 10 g L^{-1} casein, 15 g L^{-1} glycerol, 10 g L^{-1} NaCl, 100 mg L^{-1} kanamycin, 100 mg L^{-1}streptomycin, 100 mM potassium phosphate buffer (pH

7.6)].[24] When OD_{600} = 1.0, cells were cooled to 19 °C and both 20 mM sodium propionate and 100 µM IPTG were added. Cultures were incubated at 19 °C at 240 rpm for 8 d.

Measuring triketide lactone production

After 8 d, a 200-µL sample was acidified with 20 µL of 6 M HCl, centrifuged 10 min at 15,000 × g, and extracted with 1 mL ethyl acetate. Ethyl acetate was removed *in vacuo* , and 0.5 mL methanol was added. Mass spectra were obtained on an Agilent 6120 single quadrupole LC/MS, separating on a C_{18} column (5% acetonitrile: 95% water with 0.1% formic acid to 100% acetonitrile over 12 m; Agilent ZORBAX Eclipse Plus, 2.1 x 50 mm, 5 µm particle size) and scanning *m/z* 50-2000 in positive ionization mode (Data File 80). These measurements are not perfect indicators of the relative activities of **P1** -**P6** -**P7** and its point mutants. One complicating factor is that the introduced mutations could affect the expression or stability of the second protein of**P1** -**P6** -**P7** . However, SDS-PAGE analysis shows similar levels for this protein and its point mutants (Figure S6).

DATA AVAILABILITY

Data supporting the findings of this work are available within the paper and its Supplementary Information files, including Data Files 1-80 (https://github.com/adriankeatingeclay/ACPKS).

ACKNOWLEDGMENTS

This work was supported by the NIH (GM106112) and the Welch Foundation (F-1712).

AUTHOR CONTRIBUTIONS

M.H. and A.T.K performed the AlphaFold-Multimer studies. M.H. generated the **P1** -**P6** -**P7** point mutants and tested them with help from R.R.D. and S.A.. A.T.K., M.H., R.R.D., and S.A. prepared the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary Information is available for this paper. Correspondence and requests for materials should be addressed to A.T.K.

FIGURES

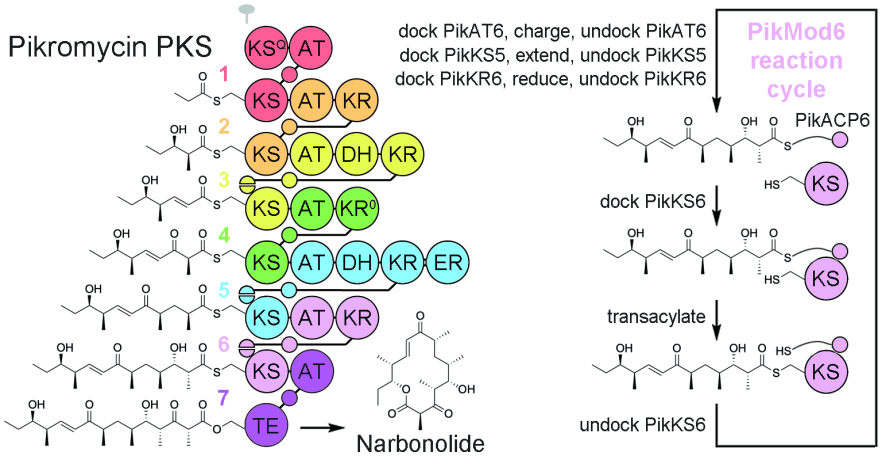

Figure 1. The docking of ketosynthase (KS) and upstream acyl carrier protein (ACP) domains within a modular polyketide synthase (PKS). The Pikromycin precursor, Narbonolide, is biosynthesized through the

actions of 23 enzymatic domains and 7 ACP domains. As each ACP carries out its reaction cycle, it docks and undocks each of its cognate enzymes. The final reaction in the reaction cycle of the first 6 modules in the Pikromycin PKS is transacylation, where an upstream ACP presents intermediates to KS and only the fully processed species is permitted onto its reactive cysteine. These checkpoints are exemplified by PikMod6, in which PikACP6 and PikKS6 associate to enable the transacylation of the hexaketide intermediate reduced by PikKR6. KS^{Q}, generates propionyl priming units; AT, acyltransferase; KR, ketoreductase; KR^{0}, epimerase; DH, dehydratase; ER, enoylreductase; TE, thioesterase; unlabeled circles, ACPs; unlabeled half-circles, docking domains.

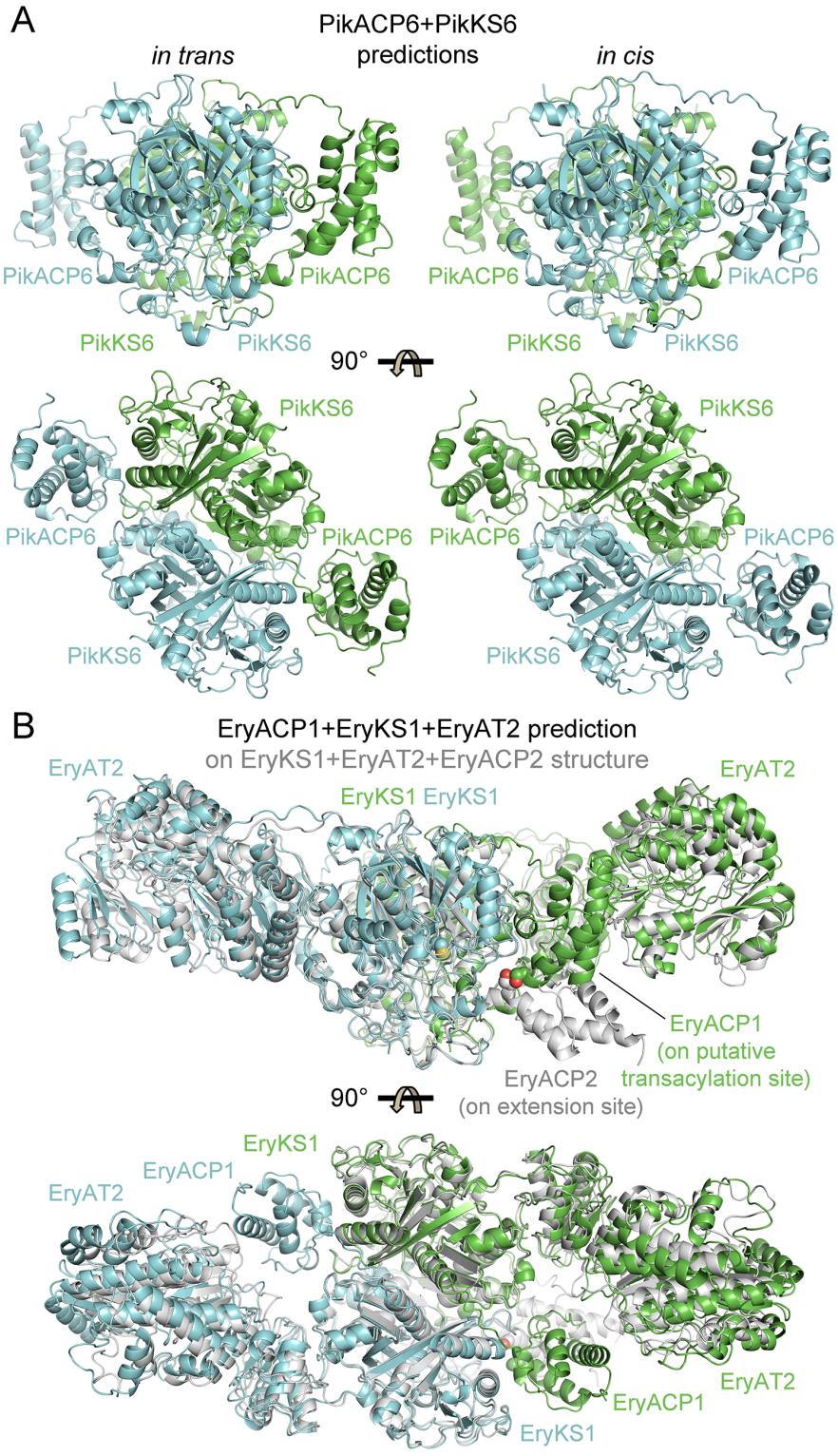

Figure 2. AlphaFold-Multimer solutions contain a putative transacylaton site. a) Two predictions for the ACP+KS dimer from PikMod6 show ACP binding to the putative transacylation site either *in trans* or

in cis . b) The top AlphaFold-Multimer prediction for the first ACP+KS+AT of the Erythromycin PKS, Ery(ACP1+KS1+AT2), superposed on the cryo-EM structure of Ery(KS1+AT2+KR2+ACP2) in which EryACP2 was observed bound at the extension site (PDB: 7M7F, EryKR2 not shown).

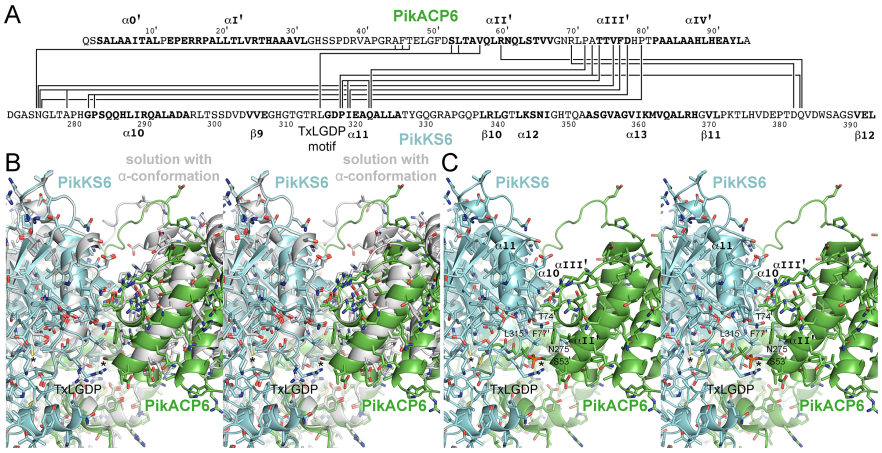

Figure 3. KS conformational change enables approach of the acyl-ACP and reactive cysteine. a) A schematic shows the contacts made in the top solution for PikMod6. b) A stereodiagram compares AlphaFold-Multimer solutions in which PikACP6 docks with PikKS6 with its TxLGDP motif in the α-conformation (gray) or the β-conformation (green and cyan). c) A stereodiagram shows the solution with the β-conformation with an acetyl-phosphopantetheinyl arm modeled between the conserved ACP serine and the KS reactive cysteine (both labeled *).

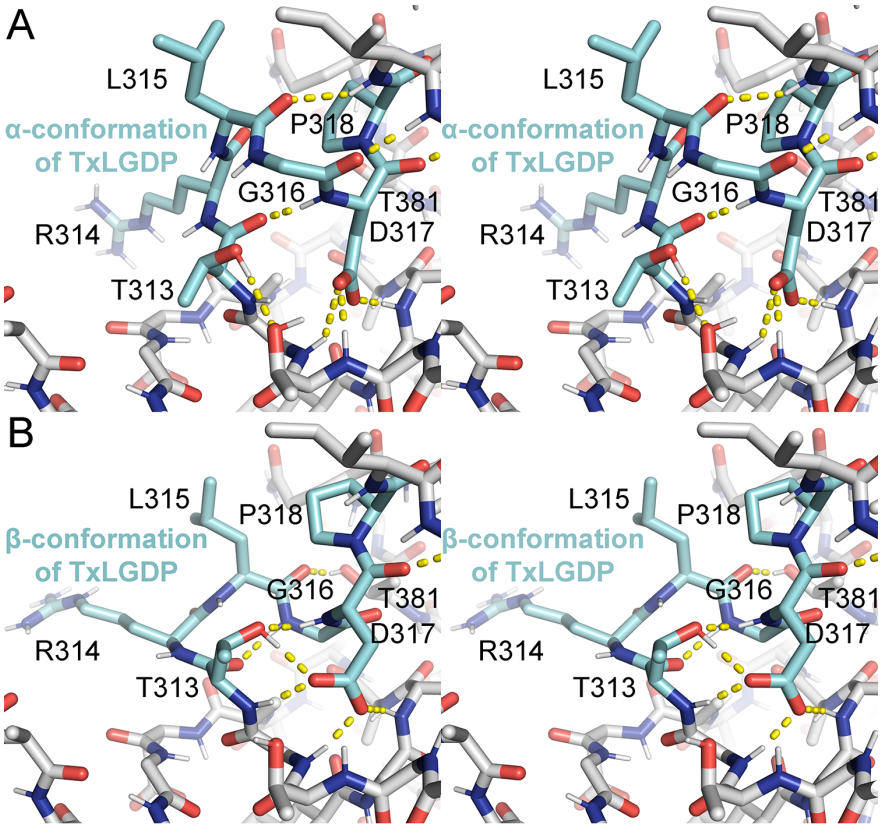

Figure 4. The α- and β-conformations of the TxLGDP motif. a) Residues 313-318 (cyan) in PikKS6 form an α-helical turn at the N-terminal end of α11, in which a hydrogen bond is formed between the T313 carbonyl and the D317 NH, as observed in all reported structures. b) ACP-bound docking solutions often show this motif (cyan) making a type I β-turn, in which hydrogen bonds are formed between the T313 carbonyl and the G316 NH as well as between the T313 hydroxyl group and the D317 NH and carboxylate. The location of the putative interface residue L315, whose C_α is shifted 4.7 Å relative to the α-conformation, is stabilized through a hydrogen bond between its carbonyl and the hydroxyl group of T381.

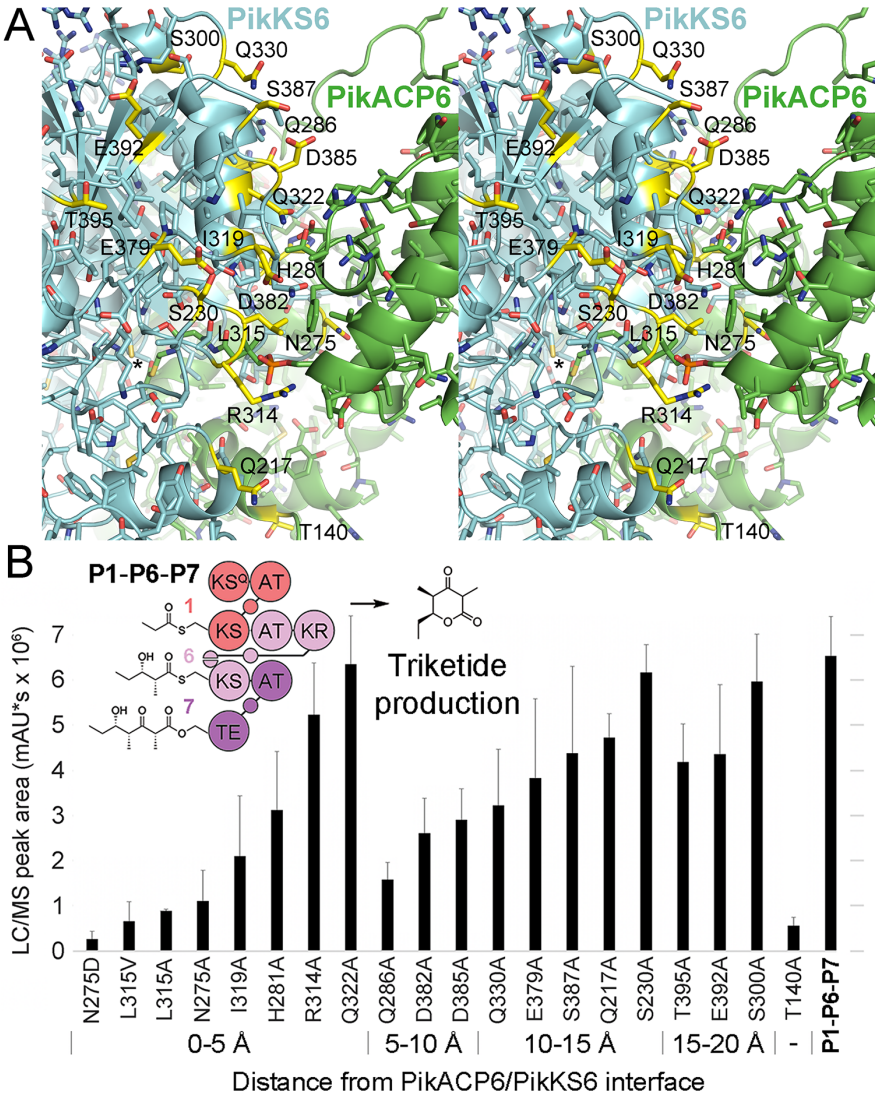

Figure 5. Mutagenesis at the proposed transacylation site. a) A stereodiagram of the top solution for PikACP6 and PikKS6 show residues in and around the interface targeted for mutational analysis (yellow). b) The model triketide lactone synthase,**P1 -P6 -P7** , was employed to measure the effects of mutating individual residues on the surface of PikKS6 (*). Triketide lactone peak areas from LC/MS analysis (triplicate experiments, standard deviations shown) were made on media extracts from 8 d cultures (Data File 80). The KS residues whose sidechains are hypothesized to make the most contact with ACP, N275 and L315, were the most sensitive to mutation. Generally, mutations to positions farther from the proposed interface resulted in smaller losses in activity.

REFERENCES

[1] Nivina, A., Yuet, K. P., Hsu, J., and Khosla, C. (2019) Evolution and Diversity of Assembly-Line Polyketide Synthases,*Chem Rev 119* , 12524-12547.

[2] Keatinge-Clay, A. T. (2017) The Uncommon Enzymology of Cis-Acyltransferase Assembly Lines, *Chem Rev 117* , 5334-5366.

[3] Helfrich, E. J. N., and Piel, J. (2016) Biosynthesis of polyketides by trans-AT polyketide synthases, *Nat*

Prod Rep33 , 231-316.

[4] Keatinge-Clay, A. T. (2012) The structures of type I polyketide synthases, *Nat Prod Rep 29* , 1050-1073.

[5] Kapur, S., Lowry, B., Yuzawa, S., Kenthirapalan, S., Chen, A. Y., Cane, D. E., and Khosla, C. (2012) Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation,*Proc Natl Acad Sci U S A 109* , 4110-4115.

[6] Kapur, S., Chen, A. Y., Cane, D. E., and Khosla, C. (2010) Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase, *Proc Natl Acad Sci U S A 107* , 22066-22071.

[7] Bagde, S. R., Mathews, II, Fromme, J. C., and Kim, C. Y. (2021) Modular polyketide synthase contains two reaction chambers that operate asynchronously, *Science 374* , 723-729.

[8] Cogan, D. P., Zhang, K., Li, X., Li, S., Pintilie, G. D., Roh, S. H., Craik, C. S., Chiu, W., and Khosla, C. (2021) Mapping the catalytic conformations of an assembly-line polyketide synthase module,*Science 374* , 729-734.

[9] Chisuga, T., Murakami, S., Miyanaga, A., Kudo, F., and Eguchi, T. (2023) Structure-Based Analysis of Transient Interactions between Ketosynthase-like Decarboxylase and Acyl Carrier Protein in a Loading Module of Modular Polyketide Synthase, *ACS Chem Biol 18* , 1398-1404.

[10] Dutta, S., Whicher, J. R., Hansen, D. A., Hale, W. A., Chemler, J. A., Congdon, G. R., Narayan, A. R., Hakansson, K., Sherman, D. H., Smith, J. L., and Skiniotis, G. (2014) Structure of a modular polyketide synthase, *Nature 510* , 512-517.

[11] Kim, S. K., Dickinson, M. S., Finer-Moore, J., Guan, Z., Kaake, R. M., Echeverria, I., Chen, J., Pulido, E. H., Sali, A., Krogan, N. J., Rosenberg, O. S., and Stroud, R. M. (2023) Structure and dynamics of the essential endogenous mycobacterial polyketide synthase Pks13, *Nat Struct Mol Biol 30* , 296-308.

[12] Desta, I. T., Porter, K. A., Xia, B., Kozakov, D., and Vajda, S. (2020) Performance and Its Limits in Rigid Body Protein-Protein Docking, *Structure 28* , 1071-1081 e1073.

[13] Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millan, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M. K., Dalwadi, U., Yip, C. K., Burke, J. E., Garcia, K. C., Grishin, N. V., Adams, P. D., Read, R. J., and Baker, D. (2021) Accurate prediction of protein structures and interactions using a three-track neural network, *Science373* , 871-876.

[14] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold,*Nature 596* , 583-589.

[15] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., and Hassabis, D. (2021) Protein complex prediction with AlphaFold-Multimer, *bioRxiv* , 2021.2010.2004.463034.

[16] Zhang, L., Hashimoto, T., Qin, B., Hashimoto, J., Kozone, I., Kawahara, T., Okada, M., Awakawa, T., Ito, T., Asakawa, Y., Ueki, M., Takahashi, S., Osada, H., Wakimoto, T., Ikeda, H., Shin-Ya, K., and Abe, I. (2017) Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides, *Angew Chem Int Ed Engl 56* , 1740-1745.

[17] Keatinge-Clay, A. T. (2017) Polyketide Synthase Modules Redefined, *Angew Chem Int Ed Engl 56* , 4658-4660.

[18] Hirsch, M., Fitzgerald, B. J., and Keatinge-Clay, A. T. (2021) How cis-Acyltransferase Assembly-Line Ketosynthases Gatekeep for Processed Polyketide Intermediates, *ACS Chem Biol 16* , 2515-2526.

[19] Vander Wood, D. A., and Keatinge-Clay, A. T. (2018) The modules of trans-acyltransferase assembly lines redefined with a central acyl carrier protein, *Proteins 86* , 664-675.

[20] Hirsch, M., Kumru, K., Desai, R. R., Fitzgerald, B. J., Miyazawa, T., Ray, K. A., Saif, N., Spears, S., and Keatinge-Clay, A. T. (2021) Insights into modular polyketide synthase loops aided by repetitive sequences, *Proteins 89* , 1099-1110.

[21] Chandran, S. S., Menzella, H. G., Carney, J. R., and Santi, D. V. (2006) Activating hybrid modular interfaces in synthetic polyketide synthases by cassette replacement of ketosynthase domains, *Chem Biol 13* , 469-474.

[22] Peng, H., Ishida, K., Sugimoto, Y., Jenke-Kodama, H., and Hertweck, C. (2019) Emulating evolutionary processes to morph aureothin-type modular polyketide synthases and associated oxygenases,*Nat Commun 10* , 3918.

[23] Miyazawa, T., Hirsch, M., Zhang, Z., and Keatinge-Clay, A. T. (2020) An in vitro platform for engineering and harnessing modular polyketide synthases, *Nat Commun 11* , 80.

[24] Miyazawa, T., Fitzgerald, B. J., and Keatinge-Clay, A. T. (2021) Preparative production of an enantiomeric pair by engineered polyketide synthases, *Chem Commun* .

[25] Zhang, J., Bista, R., Miyazawa, T., and Keatinge-Clay, A. T. (2023) Boosting titers of engineered triketide and tetraketide synthases to record levels through T7 promoter tuning, *Metab Eng 78* , 93-98.

[26] Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F., and Weissman, K. J. (2003) The structure of docking domains in modular polyketide synthases, *Chem Biol 10* , 723-731.

[27] Buchholz, T. J., Geders, T. W., Bartley, F. E., 3rd, Reynolds, K. A., Smith, J. L., and Sherman, D. H. (2009) Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains, *ACS Chem Biol 4* , 41-52.

[28] Holak, T. A., Nilges, M., Prestegard, J. H., Gronenborn, A. M., and Clore, G. M. (1988) Three-dimensional structure of acyl carrier protein in solution determined by nuclear magnetic resonance and the combined use of dynamical simulated annealing and distance geometry,*Eur J Biochem 175* , 9-15.

[29] Tang, Y., Kim, C. Y., Mathews, II, Cane, D. E., and Khosla, C. (2006) The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase, *Proc Natl Acad Sci U S A 103* , 11124-11129.

[30] Tang, Y., Chen, A. Y., Kim, C. Y., Cane, D. E., and Khosla, C. (2007) Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase, *Chem Biol14* , 931-943.

[31] Whicher, J. R., Smaga, S. S., Hansen, D. A., Brown, W. C., Gerwick, W. H., Sherman, D. H., and Smith, J. L. (2013) Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis, *Chem Biol 20* , 1340-1351.

[32] Guzman, K. M., Cogan, D. P., Brodsky, K. L., Soohoo, A. M., Li, X., Sevillano, N., Mathews, II, Nguyen, K. P., Craik, C. S., and Khosla, C. (2023) Discovery and Characterization of Antibody Probes of Module 2 of the 6-Deoxyerythronolide B Synthase, *Biochemistry 62* , 1589-1593.

[33] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr 66* , 486-501.

[34] Cunningham, B. C., and Wells, J. A. (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis, *Science 244* , 1081-1085.

[35] Murli, S., Kennedy, J., Dayem, L. C., Carney, J. R., and Kealey, J. T. (2003) Metabolic engineering of Escherichia coli for improved 6-deoxyerythronolide B production, *J Ind Microbiol Biotechnol 30* , 500-509.

[36] Sussmuth, R. D., and Mainz, A. (2017) Nonribosomal Peptide Synthesis-Principles and Prospects, *Angew Chem Int Ed Engl56* , 3770-3821.

[37] Ray, K. A. S., N.; Keatinge-Clay, A.T. (2023) Modular polyketide synthase ketosynthases collaborate with upstream dehydratases to install double bonds, *Under review at JACS* .