Jackson, D. R., Shakya, G., Patel, A. B., Mohammed, L. Y., Vasilakis, K., Wattana-Amorn, P., Valentic, T. R., Milligan, J. C., Crump, M. P., Crosby, J., & Tsai, S. C. (2018). Structural and Functional Studies of the Daunorubicin Priming Ketosynthase DpsC. *ACS Chemical Biology*, 13(1), 141-151. <https://doi.org/10.1021/acschembio.7b00551>

Peer reviewed version

License (if available):
Unspecified

Link to published version (if available):
[10.1021/acschembio.7b00551](https://doi.org/10.1021/acschembio.7b00551)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via ACS at <https://pubs.acs.org/doi/10.1021/acschembio.7b00551> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Structural and Functional Studies of the Daunorubicin Priming Ketosynthase

DpsC

David R. Jackson¹, Gaurav Shakya¹, Avinash B. Patel¹, Lina Y. Mohammed², Kostas Vasilakis², Pakorn Wattana-Amorn², Timothy R. Valentic¹, Jacob Milligan¹, Matthew P. Crump², John Crosby², and Shiou-Chuan Tsai¹

1. Department of Molecular Biology and Biochemistry, Chemistry, and Pharmaceutical Sciences, University of California, Irvine, Irvine, CA 92697.

2. School of Chemistry, University of Bristol, United Kingdom, Bristol BS8 1TS, United Kingdom.

Corresponding authors: Shiou-Chuan Tsai. Department of Molecular Biology and Biochemistry, Department of Chemistry, and Department of Pharmaceutical Sciences, University of California, Irvine. 949-824-4486. sctsai@uci.edu

John Crosby. School of Chemistry, University of Bristol, United Kingdom. +44 (0) 117 928 8445. John.Crosby@bristol.ac.uk

Matthew P. Crump. School of Chemistry, University of Bristol, United Kingdom. +44 (0) 117 331 7163. matt.crump@bristol.ac.uk

Abstract

Daunorubicin is a type II polyketide, one of a large class of polyaromatic natural products with anticancer, antibiotic and antiviral activity. Type II polyketides are formed by the assembly of malonyl-CoA building blocks, though in rare cases, biosynthesis is initiated by the incorporation of a non-malonyl derived starter unit, which adds molecular diversity to the poly- β -ketone backbone. Priming mechanisms for the transfer of novel starter units onto polyketide synthases (PKS) are still poorly understood. Daunorubicin biosynthesis incorporates a unique propionyl starter unit thought to be selected for by a sub-class ('DpsC type') of priming ketosynthases (KS III). To date, however, no structural information exists for this sub-class of KS III enzymes. Although selectivity for self-acylation with propionyl-CoA has previously been implied, we demonstrate that DpsC shows no discrimination for self-acylation or acyl-transfer to the cognate acyl carrier protein, DpsG with short acyl-CoAs. We present five crystal structures of DpsC, including *apo*-DpsC, acetyl-DpsC, propionyl-DpsC, butyryl-DpsC and a co-crystal of DpsC with a non-hydrolyzable phosphopantetheine (PPant) analogue. The DpsC crystal structures reveal the architecture of the active site, the molecular determinants for catalytic activity and homology to O-malonyl transferases, but also indicate distinct differences. These results provide a structural basis for rational engineering of starter unit selection in type II polyketide synthases.

Keywords: Daunorubicin / Doxorubicin / Polyketide / Starter Unit / Priming Ketosynthase / Anticancer

Abbreviations: PKS, KS, ACP, PPant

Introduction

Type II polyketides are a diverse group of natural products that continue to be an important source for pharmaceutical development ¹⁻⁴. Their biosynthesis is typically initiated following priming of the active site cysteine of the ketosynthase (KS) with an acetyl starter unit ⁵. Elongation of this acyl-enzyme intermediate occurs in the KS active site following decarboxylative Claisen condensations with malonate extender units similar to that seen in fatty acid biosynthesis ⁶. The potential for incorporation of alternative primers in type II polyketide synthases (PKS) creates unique opportunities to expand the chemical space of bioactive natural products ⁷ and atypical starter units were identified in daunorubicin (anticancer), R1128 (anticancer), enterocin (antibiotic), benastatin (anticancer), lomaiviticin (anticancer), and oxytetracycline (antibiotic) ⁸⁻¹⁴ (Figure 1A). The genes responsible for starter unit biosynthesis and incorporation into these type II polyketides have been identified and studied *in vitro* and *in vivo*. In the cases of enterocin, benastatin, and oxytetracycline, both experimental approaches have been applied to engineer polyketides bearing novel starter units, highlighting the importance of starter unit diversity for type II PKS bioengineering ^{11, 14-16}.

The daunorubicin PKS uses propionyl-CoA as a starter unit with the polyketide extended to a 21-carbon aromatic decaketide. Initial *in vivo* experiments appeared to rule out any involvement of the β -ketoacyl:acyl carrier protein (ACP) synthase (DpsA), chain-length determining factor (DpsB) or ACP as enzymes capable of selecting propionyl-CoA as opposed to acetyl-CoA as the biosynthetic primer ¹⁷. In subsequent experiments, however, an early propionate derived intermediate, 12-deoxyaklanonic acid, was isolated along with the acetate derived compound from a heterologous system expressing an

extended PKS complex^{8, 18}. It was proposed that fidelity of the final product could be due to the concerted activity of several enzymes, though DpsC and DpsD, homologues of the *E. coli* KAS III (EcFabH) and malonyl-CoA acyltransferase (FabD), were identified as likely to be involved in starter unit selection. Sequence comparisons suggested that DpsC could catalyze formation of 3-oxopentanoyl-S-ACP from propionyl-CoA and malonyl-CoA¹⁹, or alternatively, DpsD, could act as an acyltransferase²⁰, as it was similar to the propionyl-CoA acyltransferase domain of the 6-deoxyerythronolide B subunit from *Saccharopolyspora erythraea*^{21, 22} (Figure 1B). Expression of a gene cluster lacking *dpsC* and *dpsD* in *Streptomyces lividans* produced both a 21-carbon decaketide as well as a 20-carbon anthracycline, suggesting that neither DpsC nor DpsD were essential for the formation of aklanonic acid^{8, 18, 23} and that promiscuous starter selection can occur in their absence²⁴. When *in vivo* cell free enzyme systems were studied, however, the addition of DpsC was the major influence in starter unit selection, and therefore the synthesis of 21-carbon polyketides²⁵. Purified DpsC appeared to be monomeric and could be covalently modified at the active site serine by addition of a propionyl group, which was subsequently transferred to the phosphopantetheine thiol of ACP²⁶. DpsC was therefore identified as an acyltransferase specific for propionyl-CoA, because neither malonyl-CoA, 2-methylmalonyl-CoA nor acetyl-CoA could be transferred to the carrier protein.

The molecular basis for the substrate specificity of DpsC is unknown. Based on protein sequence homology, DpsC is similar to known priming KSs such as FabH (EcFabH) from the *E. coli* type II fatty acid synthase (FAS), and ZhuH from *Streptomyces R1128*^{6, 27}. Phylogenetic analysis, however, identified a DpsC-like KAS III sub-class due

to a predicted distinctive active site, the replacement of a usually highly conserved cysteine, present in most KSs, with a serine. DpsC also shares high sequence homology to the enzyme CerJ from *Streptomyces tendae*, which transfers malonate and malonate-derivatives onto the type II PKS natural product cervimycin ²⁸.

DpsC is an attractive target for understanding and engineering PKSs because of its unique dual-role activities and ability to incorporate an atypical starter unit in biosynthesis. Here we re-examine the role of DpsC by investigating the reported selectivity for self-acylation with propionyl-CoA. We also present the high-resolution X-ray structure of *apo*-DpsC that represents the first example of the priming active site of this class of KAS III enzymes. Three high-resolution crystal structures of acyl-enzymes including acetyl-, propionyl- and butyryl-DpsC have also been solved. Analysis was then extended with a fourth co-crystal structure, which included an acyl-pantetheine analogue that mimics the pre-acylation complex. We employ knowledge of the DpsC active site to identify homolog enzymes with verified or putative AT activity, and use homology modeling to rationalize the observed substrate scope. Together, these results provide a structural basis for starter unit engineering in type II PKSs and expand our understanding of AT and KS mechanisms.

Results and Discussion

Self-Acylation of DpsC. Previous *in vitro* studies showed that DpsC can self-load propionate from propionyl-CoA and selectivity for this substrate was implied^{25, 26}. We examined self-acylation of DpsC with acetyl-CoA, propionyl-CoA, malonyl-CoA and butyryl-CoA, and observed that DpsC loaded each of these substrates as characterized by ESI-MS (Figure S1). An initial step-wise assay monitored by ESI-MS also demonstrated that the self-acylation reaction went to between 80-100% completion. To determine kinetic parameters for the self-acylation of DpsC, a continuous coupled enzyme assay was employed²⁹⁻³². (Figure S2). Acetyl-, propionyl- and butyryl-CoA were equally good substrates for DpsC with very similar values of K_M , maximum velocity (V_{max}) and k_{cat}/K_M . The K_M for malonyl-CoA on the other hand was lower. In addition, the lower V_{max} and turnover number (k_{cat}) for this acyl coenzyme indicates that malonate is processed at a considerably slower rate. The k_{cat}/K_M ratio was nearly identical for acetyl, propionyl, and butyryl species. Loading of malonate onto DpsC was, however, faster overall presumably due to the fact that malonyl-CoA is very effective at saturating the active site of DpsC. Neither methylmalonate nor hexanoate were transferred from their respective acyl-CoAs to DpsC indicating that the active site architecture of DpsC can discriminate against branched or elongated CoA substrates.

The acyltransferase activity of DpsC was then investigated by incubating DpsG and DpsC with either acetyl-, propionyl- or butyryl-CoA and transfer monitored by ESI-MS (Figure S3A). In each case, DpsC was able to transfer acetate, propionate and butyrate onto DpsG. In the absence of DpsC, *holo*-DpsG was unable to self-load any of the priming units (data not shown). The acyltransferase activity of DpsC was further

investigated by examining the rate at which DpsC converted *holo*-DpsG to the acylated form. To achieve this, samples from the acyltransferase assay were analyzed by ESI MS. The rate at which DpsC converted DpsG to the acylated form was similar for all starter units and in contrast to previous reports that suggested that DpsC was only capable of transferring propionate onto the ACP (Figure S3B) ²⁶. As with self-loading, malonyl transfer to *holo*-DpsG was observable but slow and only reached 20% completion. Finally, we tested the ability of DpsC to selectively transfer propionate onto *holo*-DpsG from a pool of different starter units. When DpsC and DpsG were incubated with a mixture of equimolar concentrations of acetyl-CoA, propionyl-CoA and butyryl-CoA, it was observed that DpsC would transfer all three starter units to DpsG (Figure S3C).

In summary, contrary to previous suggestions that DpsC may be important for starter unit selection ^{25, 26}, we found that the substrate specificity of acyl-CoAs are highly similar for DpsC. Transfer of acyl-CoAs to the carrier protein DpsG also did not introduce any degree of selection, whether incubated with a single acyl-CoA species or a selection of low molecular weight acyl-CoAs. This shows that *in vitro* propionyl-CoA specificity is not determined by DpsC or the transfer of the acyl group to the phosphopantetheinyl thiol of the ACP.

Overall Structure of DpsC. Recombinantly expressed His₆-tagged DpsC was obtained following affinity chromatography to leave the purified protein (ESI-MS – observed 39243; calculated 39248 Da). Gel filtration and native gel electrophoresis both showed that DpsC had a molecular mass of 90 kDa, consistent with a dimeric structure. *Apo*-DpsC crystallized in the space group P6₅22 with two DpsC molecules per asymmetric unit. The

structure was solved to a resolution of 2.5 Å and revealed a thiolase fold common to ketosynthases and the acyltransferase CerJ ²⁸ (Figure 2A-B, Figure 3A-D). Two DpsC monomers form a homodimer, with each monomer related by a 2-fold rotational axis of non-crystallographic symmetry. The DpsC dimer interface area is estimated to be 2740 Å² (PISA) ³³, which is similar to the *EcFabH* interface ⁶. When overlaid, the DpsC monomers are almost identical with an overall backbone RMSD of 0.18 Å. The secondary structure of each DpsC monomer comprises two consecutive β-α-β-α-β-α-β motifs (termed the N- and C-terminal cores) connected by a series of loops, several of which contain additional secondary structure elements. The N-terminal core includes Nα1-Nα3, Nβ1-Nβ5, and loops L1-L8. L1 contains two α-helices and two β-sheets designated L1(α1), L1(α2), L1(β1), and L1(β2). The C-terminal core consists of Cα1-Cα3, Cβ1-Cβ4, and loops L9-L15. L9 contains two α-helices designated L9(α1), and L9(α2), while L12 includes one α-helix designated L12(α1). In DpsC, the missing β-strand in the second core motif is a small loop between Cα2 and the L12(α1) region (Figure 2C-D).

Features of the DpsC Active Site that Enable AT Activity. The active site of DpsC contains an unusual Cys to Ser118 substitution (Figure 3) ^{6, 23, 27, 34} In *EcFabH* a Cys to Ser active site mutation eliminated transacylation ⁶, whereas DpsC has retained this essential function. Typically, in the priming KSs, the active site cysteine is situated on a bend between L5 and the N-terminal of Nα3 known as the ‘nucleophilic elbow’. The cysteine residue is acylated with an acyl-CoA derived starter unit and histidine and asparagine facilitate decarboxylative condensation of this acyl group with malonyl-ACP ⁶. In DpsC, Ser118 replaces the usual KS Cys, though its positioning at this juncture is

conserved (Figure S4A and Figure 3E). The precise location of the active site nucleophile is critical as it gains nucleophilicity through lowering the pKa of the ionisable –SH or –OH group via stabilization from the macroscopic helix-dipole effect^{35, 36}. His244 and Asn274 make up the remainder of the catalytic triad in *EcFabH* (PDB ID: 1EBL) where these residues are located on loop L11 and loop L12, respectively (Figure S4B). The two residues in DpsC that most closely match their three-dimensional arrangement are Asp302 and His297 and they appear to complete this catalytic triad although they differ in being located on helix Cα3 and loop L12, respectively (Figure S4A). DpsC His297 occupies the same relative position to the nucleophilic Ser/Cys as *EcFabH* Asn274 but is positioned at the back of the active site pocket below the catalytic Ser118 and is less accessible to the substrates when compared to Asn274 of *EcFabH* (Figure S4C-D). The imidazole ring of His297 is also rotated by 90° relative to the *EcFabH* Asn274 side chain amide. This brings Nε2 of His297 within hydrogen bonding distance of the catalytic Ser118, which may increase the nucleophilicity of this residue (Figure 3E, S4C-D). The alternate orientation of the Asn274 amide group of *EcFabH* does not permit an equivalent hydrogen bond with Cys112. Similar arguments also apply to the polyketide priming KSs such as ZhuH²⁷ (PDB ID: 1MZJ) (Figure 3E). The remaining imidazole nitrogen, ND1, is oriented correctly to form a hydrogen bond with Asp302 on helix Cα3 (Figure S4A). Asp302 may further enhance the nucleophilicity of Ser118 by polarizing His297. Therefore, the Ser-His-Asp catalytic triad of DpsC shows obvious differences to the catalytic triad of priming FAS and PKS KSs^{6, 27} and the direct contact between His297 and S118 is more closely related to serine protease Ser-His-Asp catalytic triads or acyltransferase (ATs) Ser-His catalytic dyads than typical KS active site triads (Cys-His-

Asn or Cys-His-His)^{37, 38}. DpsC has 41% sequence identity with the recently characterized O-malonyl transferase CerJ where the active site consists of a Cys-His-Asp catalytic triad. These residues, except for the Cys/Ser switch, are positioned on the same secondary structural elements and in the same orientation (Figure 3E) to those of DpsC.

Acetyl-, Propionyl-, and Butyryl-DpsC Crystal Structures. To further understand how DpsC recognizes acyl-CoAs, we solved three co-crystal structures of DpsC bound with different acyl groups. *Apo*-DpsC crystals were soaked with acetyl-CoA, propionyl-CoA, and butyryl-CoA to generate the respective covalently bound acyl-enzymes (Figure 4 and Figure S5). These structures overlay with *apo*-DpsC with overall RMSDs of less than 0.18 Å. In each the L9 loop has elevated B-factors for both monomers and residues Gln212-Arg213 in the *apo*-DpsC structure and residue Arg213 in the L9 loop region of acetyl-DpsC could not be modeled because of a lack of electron density. The active sites of acetyl-, propionyl-, and butyryl-DpsC are all very similar but there are, however, discernible changes to the active sites when compared to *apo*-DpsC. In the *apo*-DpsC structure the active site Ser118 hydroxyl oxygen is 2.7 Å from Nε2 of His297. In the acetyl-, propionyl-, and butyryl-DpsC enzymes, this distance increases to 3.2 Å, 3.1 Å, and 3.1 Å, respectively, as the Cβ-Oγ bond rotates away and the ester carbonyl group occupies an oxyanion hole formed by the amide protons of Ser118 and Ala329 (Figure 4). In the *apo*-DpsC structure, however, water fills the equivalent space. The conserved binding of the carbonyl group directs the rest of the substrate into the same region of the binding pocket and is enclosed by Leu86, Leu93, and Val117. Collectively, these residues define the pocket size. Leu93, contributed by loop L3 of the opposing subunit of the dimer caps

off the cavity at the tip of the longer acylated species (Figure 4) and may be an important structural element in defining the upper size limit of bound acyl-substrates. In *EcFabH*, Phe87 has been attributed to restricting the size of the incoming substrate and Leu93 of DpsC overlaps well with Phe87 of *EcFabH* (Figure S6) ³⁹. DpsC is therefore not capable of specific self-acylation selection between short chain (< 4 carbons) unbranched CoAs as has been previously suggested. The ability of DpsC to accept multiple acyl-CoAs as substrates for self-acylation suggests that DpsC is promiscuous but it may be restricted to a limited pool of small, non-branched acyl substrates.

DpsC Co-crystal Structure with a Substrate Analogue. We developed a propionyl-amide pantetheine analogue (**4**, Figure 5A) to capture an acyl substrate-DpsC complex just prior to transfer to DpsC (Figure 5). **4** was synthesized by coupling a protected pantetheine analogue (**1**) with propanoic acid followed by deprotection to produce **3** (Figure S7). The terminal hydroxyl of **3** was then chemoenzymatically phosphorylated using the enzyme CoaA to yield **4**. This compound was co-crystallized by soaking with DpsC (Figure 5). Electron density of probe **4** is observed within the substrate tunnel of both monomers, although the electron density for **4** is better defined in monomer A (Figure 5B). Lys221 (L9- α 2) and Arg271 (Ca2) interact strongly with the terminal phosphate of **4**. This type of interaction has also been observed in structures of other PKS enzymes where PPant is bound, demonstrating the importance of the PPant phosphate for binding (Figure 5C-D). In the crystal structure of the *beta*-branching enzyme CurD from the curacin biosynthetic pathway in complex with an acetyl-donor ACP (PDB ID: 5KP8), for example, the terminal phosphate of the ACP-bound PPant moiety interacts strongly with two

arginine side chains ⁴⁰. Moving towards the active site, the dimethyl moiety of **4** binds a hydrophobic pocket between Ca1, L9- α 2, and L9- α 1. This trend is also observed in other crystal structures, including the CurD-ACP structure as well as the crystal structure of the AT enzyme VinK from the vicenistatin biosynthetic pathway in complex with a cognate ACP (PDB ID: 5CZD) ⁴¹. Given their locations near the surface of these enzymes and their prevalence in published structures, the PPant phosphate-binding site and hydrophobic pocket are possibly two common features of PKS enzymes that serve to bind and orient the PPant group of an incoming acyl-CoA or ACP ⁴¹. The hydroxyl group of **4** has no direct hydrogen bonds with DpsC; however, a water molecule is coordinated between this hydroxyl group, the terminal phosphate and an amide proton of **4** (Figure 5E). The two central carbonyl groups of **4** interact with a second water molecule with no other significant interactions with DpsC. Two of the methylene groups of **4** also form hydrophobic interactions with Val327 (C β 3). Additionally, the carbonyl group of the propionyl amide bond binds within the active site via the backbone NH of Ala329. This type of interaction was also observed between the acyl carbonyl and two backbone amides deep within the active site of CurD in the CurD-ACP structure ⁴⁰. Finally, Ser118 is poised for acyl transfer, with the electrophilic carbon of the carbonyl group 3.0 Å away from the activated hydroxyl nucleophile and the approach angle close to the ideal Bürgi-Dunitz angle (107°) for nucleophilic attack on a trigonal carbonyl centre (Figure 5D). The amide carbonyl interactions are similar to those observed for the acyl-DpsC structures. Likewise, the hydrophobic propionyl moiety binds a hydrophobic cleft formed by Leu217, Leu93, Leu86, and Val117. When overlaid with the propionyl-DpsC structure, the DpsC co-crystal structure with **4** shows that the propionyl group attached to Ser118 is in the

same orientation as the propionyl group of **4** (Figure 5F). In summary, the DpsC co-crystal structure shows that **4**, with the full-length PPant group, extends deep into the active site and does not induce major conformational changes when compared to the *apo*- or acyl-DpsC structures and shares a similar binding mode to other PPant binding enzymes.

Identification of DpsC substrate binding residues rationalizes the substrate scope of other ATs with KS folds. The active site of DpsC contains an atypical Ser-His-Asp catalytic triad, which has elements of both AT and KS architectures. A protein BLAST search using the amino acid sequence of DpsC revealed many uncharacterized bacterial enzymes annotated as KS III-like, with the AT, CerJ, sharing the highest sequence identity. A series of other putative KSs were identified along with the natural product produced from their biosynthetic gene clusters and aligned (Figure 6 and Figure S8). Most of these KS enzymes are from gene clusters whose products contain acylated glycosides. In many cases, acylation appears selective for the hydroxyl at the 4 position of terminal hexose moieties (cervimycins, polyketomycin, avilamycin A, and SF2575). The DpsC homologues in these gene clusters may use acyl-CoAs as substrates to acylate glycosides similar to CerJ; therefore, their active sites are expected to have residues and sizes that reflect their corresponding substrate specificity. Homology models of these DpsC homologues identified by BLAST were generated and overlaid with the DpsC structure to determine their putative active site residues regarding substrate binding.

The table describing residues surrounding the active sites of the homology models is detailed in Figure 7A. The size of the putative substrates correlate with the size of the substrate pockets as reflected by the decreasing size of the side chains of residues that

define this pocket (Figures 6-7, S8). Residues corresponding to Leu93, termed the active site cap of DpsC, for example, change to glycine as the putative substrate size increases, which might expand the available volume within the acyl-binding pocket (Figure 7A). There is also a switch in the middle pocket residue corresponding to Leu217 of DpsC as the substrate goes from aliphatic to aromatic. Residue Leu217 of DpsC is located at the side of the active site pocket, and changes to an aromatic residue, phenylalanine or tyrosine, if the putative substrate is aromatic. This trend at the Leu217 position may occur because of favorable π - π stacking between phenylalanine or tyrosine and an aromatic substrate. Finally, residues at the top and bottom of the active site pocket, corresponding to Val117 and Tyr165 of DpsC, also change to smaller residues as the substrate becomes larger (Figure 7). The analysis of DpsC homologues with different substrate specificities helps determine residues important for substrate specificity and reveals insights into how the DpsC active site could be engineered to introduce new biosynthetic starter units. Moreover, identification and determination of the substrate tolerance of enzymes that regioselectively modify glycosyl moieties of natural products may provide powerful opportunities to site-specifically modify existing natural products and diversify known bioactive scaffolds.

Conclusion.

The lack of affinity for starter units other than propionate led Hutchinson and co-workers to suggest that DpsC was in fact the starter unit fidelity factor and led us to investigate this property of DpsC further ²⁶. In this study however we observed non-specific activity of DpsC which was able to self-load acetyl-, propionyl- and butyryl-CoA

and then transfer these onto DpsG. Therefore neither DpsC nor DpsG encode this important activity.

Five DpsC crystal structures are reported in this paper: *apo*-DpsC, three acyl-DpsC structures, and a DpsC co-crystal structure with the substrate analogue **4**. These structures reveal that the catalytic triad (Ser-His-Asp) of DpsC differs from the active site triads of other KS IIIs (Cys-His-Asn). The geometry of the active site resembles that of the acyltransferase CerJ, a first observation for a type II priming KS. All three acyl-DpsC structures showed the acyl group bound in a hydrophobic cleft at the bottom of the substrate-binding pocket, with the ester carbonyl group oriented towards an oxyanion hole formed by the amide protons of Ala329 and Val117. The Ser118 hydroxyl forms a hydrogen bond with His297, whose developing dipole is then stabilized by Asp302. This increases the nucleophilicity of the serine oxygen and explains the acyltransferase activity of DpsC. The co-crystal structure with **4** revealed important binding interactions between the PPant group and the enzyme. This indicated the orientation of the substrate within the acyl-binding region and corroborates the assay data which shows that propionate-specificity is not defined by DpsC. We are now investigating the potential roles of other enzymes, including the acyltransferase DpsD, which may collectively orchestrate propionate selection. It is likely that the protein-protein interactions among DpsC, DpsD and DpsG determines the starter unit preference, and the interplay between DpsC-DpsD-DpsG is currently under investigation. Finally, bioinformatic analysis highlighted a suite of putative ATs containing the KS fold, which can be further exploited for engineered biosynthesis of polyketides with novel starter units or modifications of glycosides.

Materials and Methods

Expression and Purification of DpsC. The pET-28b(+) (Novagen) derived DNA plasmid encoding N-terminal His-tagged DpsC (DpsC/pET28b) was transformed into *E. coli* BL21(DE3) competent cells and plated on LB-agar plates containing kanamycin (50 µg/mL). The plates were incubated overnight at 37 °C. Positive transformants were transferred to a 5 mL starter culture of Luria-Bertani (LB) broth containing kanamycin (50 µg/mL) and grown overnight at 37 °C with shaking, which was then used to inoculate one liter of LB with kanamycin (50 µg/mL). Cultures were grown at 37 °C until the A_{600} reached 0.4 – 0.6. The cells were then cooled to 18 °C, and 0.1 mM IPTG was added to induce protein expression. After 12-18 hours of incubation at 18 °C, the cells were harvested by centrifugation at 5000 rpm for 10 minutes. The cell pellets were flash-frozen in liquid nitrogen and stored at -80 °C. The frozen cell pellets were thawed on ice and resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 % glycerol, 10 mM imidazole). The cell suspension was lysed using sonication (8 x 30 s cycles), and cellular debris was removed by centrifugation at 14000 rpm for 45 minutes. The lysate was incubated with 5 mL Ni-IMAC resin (BioRad) at 4 °C for one hour. The resin was poured into a fritted column and the flow-through fraction was collected. The resin was washed with 100 mL of lysis buffer then eluted with lysis buffer plus increasing amounts of imidazole (20-500 mM). The elutions were analyzed using SDS-PAGE, and the elutions containing the protein of interest were combined and dialyzed overnight into storage buffer (50 mM Tris pH 8.0, 100 mM NaCl, 10 % glycerol). 1 unit of bovine thrombin (Sigma) per mg of DpsC was added to the combined protein elutions prior to dialysis to cleave the N-terminal His-

tag. Following the dialysis, DpsC was incubated with Ni-IMAC resin and passed over a fritted column to remove any uncleaved DpsC and the flow through was analyzed by MALDI-TOF MS to confirm cleavage of the His-tag. The cleaved DpsC protein was then concentrated to 5 mg/ml, buffer exchanged into crystallization buffer (25 mM Tris, pH 7.5), and flash frozen in liquid nitrogen.

Expression and Purification of DpsG.

To ensure production of *holo*-ACP, *DpsG* and *ACPS* were cloned as adjacent genes into pET15b and used to transform *E. coli* BL21(DE3) cells. Cells were grown and ACP purified as reported previously with the addition of a single gel filtration step ⁴². ESI-MS gave an observed mass for DpsG of 9604 Da (calculated 9602 with loss of the N-terminal methionine).

Self-acylation and transfer assays.

Self-acylation of DpsC with acyl-CoA. 40 μ M DpsC was incubated with 200-800 μ M propionyl-CoA or other CoA-derivatives in the presence of assay buffer containing 2 mM TCEP in a 100 μ L reaction volume. The reaction was incubated at 30 °C for 2 hours. The assay was then concentrated using C4 resin prior to ESI-MS analysis. The ESI-MS spectra were recorded using a Fisons VG Quattro triple quadrupole or QSTAR XL hybrid quadrupole time-of flight mass spectrometer. For acyltransfer to DpsG, the 100 μ L assay consisted of 50 μ M DpsG, 1 μ M DpsC, 1 mM CoA-derivatives and 2 mM TCEP in assay buffer was incubated at 30°C for 2 hours. The assay was then concentrated using C4 resin prior to ESI-MS analysis

Kinetic assessment of self-acylation. To determine kinetic parameters for the self-acylation of DpsC, a continuous coupled enzyme assay was employed as previously described³². These assays couple the release of free coenzyme A generated by the self-acylation of DpsC to the oxidative decarboxylation of α -ketoglutaric acid (α KG), catalyzed by the α -ketoglutarate dehydrogenase (KDH) enzyme complex. The free CoASH generated is used as a substrate by KDH to produce succinyl CoA with concomitant reduction of NAD^+ to NADH that can be followed spectrophotometrically at 340 nm ($\epsilon = 6220 \text{ M}^{-1} \text{ cm}^{-1}$). To ensure that the coupling enzyme was in its *holo*- form, KDH was incubated with α KG, NAD^+ , TPP and either acetyl-, propionyl-, butyryl- or malonyl-CoA for 5 min prior to the addition of DpsC. The latter was also preincubated for 5 minutes at the reaction temperature. DpsC was then added to the mixture of the other components and initial reaction rates (5-10% of the reaction) were measured for varying concentrations of CoA derivatives by monitoring the linear regression ($V_0 = d[\text{NADH}]/dt$) and verifying that the self-acylation reactions obeyed Michaelis-Menten kinetics. Kinetic parameters were calculated from the hyperbolic regression of the initial rates of self-acylation against varying substrate concentrations.

Crystallization of DpsC, Acetyl-DpsC, Propionyl-DpsC, Butyryl-DpsC, and DpsC + PPant Analogue 4. Frozen aliquots of DpsC were thawed on ice then filtered with a 0.22 μM filter prior to crystallization. DpsC was crystallized using the sitting drop vapor diffusion method at room temperature from crystal seed stocks generated using a Seed Bead (Hampton). Crystals were grown by adding one microliter of 5 mg/mL protein solution to

one microliter of well solution (0.18 M sodium citrate, 26 % PEG 3350) containing crystal seeds and allowed to equilibrate with 500 μ L of well solution. Three-dimensional hexagonal crystals formed over one week. For acetyl-DpsC, propionyl-DpsC, and butyryl-DpsC, seeded wild type DpsC crystals were soaked with 5 mM acetyl-CoA, 5 mM butyryl-CoA, or 5 mM propionyl-CoA 12-16 hours before being frozen in liquid nitrogen. For co-crystallization with **4**, DpsC was incubated on ice for one hour with 5 mM of **4**, and diffraction quality crystallization was achieved with multiple rounds of seeding, starting with wild type DpsC crystals and subsequent rounds with co-crystal seeds. X-ray diffraction data was collected for apo-DpsC and all acyl-DpsC crystals at the Advanced Light Source (ALS) using either Beamline 8.2.1 or 8.2.2. The resolutions of the data sets were 2.50 Å, 2.40 Å, 2.40 Å, and 2.20 Å, for apo-, acetyl-, propionyl-, and butyryl-DpsC, respectively. Diffraction data for the DpsC-propionyl pantetheine analogue **4** co-crystals was collected at the Stanford Synchrotron Radiation Laboratory (SSRL), Beamline 12-2, to a resolution of 2.30 Å. All data were processed using indexed, integrated, and scaled using HKL2000 ⁴³. Crystallographic data and refinement statistics are presented in Table S1.

DpsC Phasing, Model Building, and Refinement. DpsC crystallized in the space group P6₅22 with two DpsC monomers forming a dimer in the asymmetric unit. The *apo*-DpsC structure was solved using molecular replacement (Phaser) with CerJ (PDB ID: 3S3L) as a starting model ⁴⁴. A preliminary model was built using (PHENIX AutoBuild), which was then used for iterative rounds of model building (COOT) and refinement (PHENIX Refine) ⁴⁵⁻⁴⁷. The model was refined to an R_{work} of 20.9% and R_{free} of 16.2%. Acetyl-DpsC,

propionyl-DpsC, butyryl-DpsC, and DpsC co-crystal with **4** were all solved by molecular replacement using the *apo*-DpsC structure as a search model. Model building and refinement were carried out in a similar manner as described for DpsC above with a few exceptions. For acetyl-, propionyl-, and butyryl-DpsC, after multiple rounds of refinement, there was well-defined electron density for the acyl groups attached to Ser118. The acyl-serine moiety was added using COOT and further refined. For the DpsC co-crystal with **4**, multiple rounds of refinement generated well-defined electron density for **4**. The PDB file for the ligand was built using the NCI SMILES converter server (<http://cactus.nci.nih.gov/translate/>). The ligand was manually docked in the DpsC active site using COOT with restraints generated using PHENIX eLBOW ⁴⁸. Data collection and refinement statistics for all structures can be found in Table S1.

Synthesis of Compound 4. ¹H NMR spectra were recorded at 298 K on a Bruker 400 MHz spectrometer. ¹³C NMR spectra were recorded at 298 K on a Bruker 125 MHz spectrometer. ¹H and ¹³C spectra were referenced to residual chloroform (7.26 ppm, ¹H; 77.00 ppm, ¹³C) or residual methanol (3.31 ppm, ¹H; 49.00, ppm ¹³C). The raw fid files were processed using XWinNMR (Bruker Corporation). The two-step synthesis is detailed in Figure S7.

Coupling of Protected-pantethenamine 1 with propionic acid. EDAC•HCl (385 mg, 2.011 mmol) was added to a mixture of propionic acid (60 μ L, 0.80 mmol) and amine (protected-pantethenamine **1**) (305 mg, 0.80 mmol) dissolved in dry DMF (5 mL). Compound **1** was synthesized as previously described ⁴⁹. Within 5 min, EtNⁱPr₂ (0.7 mL, 4 mmol) was added

and the mixture was stirred at rt. After 12 h at rt, the solvent was removed by rotary evaporation. Flash chromatography (EtOAc to 3:1 EtOAc:MeOH) afforded 204 mg (58%) of amide **2** as colorless oil. **¹H NMR** (400 MHz, CDCl₃) δ 7.42 (d, *J* = 8.6 Hz, 2H), 7.03-7.00 (m, 1H), 6.93-6.90 (d, *J* = 8.7 Hz, 2H), 6.74-6.72 (m, 1H), 6.32-6.30 (m, 1H), 5.46 (s, 1H), 4.07 (s, 3H), 3.82 (s, 3H), 3.68 (AB, 2H), 3.56-3.50 (m, 2H), 3.37-3.28 (m, 4H), 2.42 (t, *J* = 6.3 Hz, 2H), 2.16 (q, *J* = 7.6 Hz, 2H), 1.12 (t, *J* = 7.6 Hz, 3H), 1.10 (s, 3H), 1.09 (s, 3H) (Figure S9). **¹³C NMR** (125 MHz, CDCl₃) δ 175.0, 172.0, 169.6, 160.3, 130.2, 127.6, 113.8, 101.4, 83.9, 78.5, 55.4, 40.1, 36.2, 34.9, 33.1, 29.6, 21.9, 19.2, 9.9 (Figure S10).

Deprotection step. The protected amide **2** obtained from previous step (68 mg, 0.156 mmol) was dissolved in 80% aq. AcOH (2 mL) open to air. After 16 h at rt, the solvent was removed via rotary evaporation assisted by azeotropic removal of water with toluene. Flash chromatography (EtOAc to 4:1 EtOAc:MeOH) afforded 45.5 mg (92%) of desired product **3** as a colorless oil. Spectroscopic data are in agreement with those previously published⁵⁰. **¹H NMR** (400 MHz, CD₃OD) δ 3.90 (s, 1H), 3.53-3.45 (m, 4H), 3.28 (m, 4H), 2.43 (t, *J* = 6.6 Hz, 2H), 2.22 (q, *J* = 7.6 Hz, 2H), 1.14 (t, *J* = 7.6 Hz, 3H), 0.93 (s, 6H) (Figure S11). **¹³C NMR** (125 MHz, CD₃OD) δ 177.4, 176.1, 174.2, 77.3, 70.3, 40.4, 40.0, 36.6, 36.4, 30.2, 21.3, 20.9, 10.4 (Figure S12).

Chemoenzymatic Phosphorylation of 3. Compound **3** was incubated with the pantetheine kinase CoaA (5 μM) in the phosphorylation buffer (25 mM potassium phosphate, 10 mM MgCl₂, 8 mM ATP, pH 7.5) for 3 hours at 37 °C to generate **4**. The final product **4** was purified by HPLC using a semi-prep C18 column with a linear gradient

of water/acetonitrile containing 0.1 % (v/v) formic acid. Fractions containing **4** were collected and verified by ESI-MS then lyophilized, resuspended in water, and stored at -20 °C.

Sequence Alignments and the Generation of Homology Models of DpsC

Homologues. A search for proteins similar to DpsC was conducted using BLAST that searched for non-redundant protein sequences. Genes from known biosynthetic gene clusters were analyzed for products containing an acylated sugar moiety, and they were chosen for homology model generation using HHPred⁵¹. Sequence alignments were conducted using ClustalW and graphics were generated using Esript (<http://esript.ibcp.fr>).

Supporting Information. A pdf file of Supporting Information, which contains crystallographic statics, compound NMR spectra, and additional protein models, is available free of charge on the ACS Publications website at <http://pubs.acs.org>.

ACKNOWLEDGMENT

This work was funded by NIH GM100305 and GM076330 (to S.-C.T.). We also wish to thank the Royal Thai Government for DPST Fellowship for P. W.-a, the Human Capacity Development Program (HCDP) from Ministry of Higher Education and Scientific Research / Iraq- Kurdistan region for L. Y. M. and E.P.S.R.C. for a doctoral training award to K.V. This work was part of the DOE Joint BioEnergy Institute supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental

Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393).

References

1. Cragg, G. M., and Newman, D. J. (2001) Natural product drug discovery in the next millennium, *Pharm. Biol.* 39S, 8-17.
2. Zhang, M. Q., and Wilkinson, B. (2007) Drug discovery beyond the 'rule-of-five', *Curr Opin. Biotechnol.* 18, 478-488.
3. Harvey, A. L. (2008) Natural products in drug discovery, *Drug Discov. Today* 13, 894-901.
4. Weber, T., Welzel, K., Pelzer, S., Vente, A., and Wohlleben, W. (2003) Exploiting the genetic potential of polyketide producing streptomycetes, *J. Biotechnol.* 106, 221-232.
5. Hertweck, C., Luzhetskyy, A., Rebets, Y., and Bechthold, A. (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork, *Nat. Prod. Rep.* 24, 162-190.
6. Davies, C., Heath, R. J., White, S. W., and Rock, C. O. (2000) The 1.8 Å crystal structure and active-site architecture of beta-ketoacyl-acyl carrier protein synthase III (FabH) from escherichia coli, *Structure* 8, 185-195.
7. Moore, B. S., and Hertweck, C. (2002) Biosynthesis and attachment of novel bacterial polyketide synthase starter units, *Nat. Prod. Rep.* 19, 70-99.
8. Rajgarhia, V. B., and Strohl, W. R. (1997) Minimal Streptomyces sp. strain C5 daunorubicin polyketide biosynthesis genes required for aklanonic acid biosynthesis, *J. Bacteriol.* 179, 2690-2696.
9. Piel, J., Hertweck, C., Shipley, P. R., Hunt, D. M., Newman, M. S., and Moore, B. S. (2000) Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate 'Streptomyces maritimus': evidence for the derailment of an aromatic polyketide synthase, *Chem. Biol.* 7, 943-955.
10. Marti, T., Hu, Z., Pohl, N. L., Shah, A. N., and Khosla, C. (2000) Cloning, nucleotide sequence, and heterologous expression of the biosynthetic gene cluster for R1128, a non-steroidal estrogen receptor antagonist. Insights into an unusual priming mechanism, *J. Biol. Chem.* 275, 33443-33448.
11. Xu, Z., Schenk, A., and Hertweck, C. (2007) Molecular analysis of the benastatin biosynthetic pathway and genetic engineering of altered fatty acid-polyketide hybrids, *J. Am. Chem. Soc.* 129, 6022-6030.
12. Kersten, R. D., Lane, A. L., Nett, M., Richter, T. K., Duggan, B. M., Dorrestein, P. C., and Moore, B. S. (2013) Bioactivity-guided genome mining reveals the lomaiviticin biosynthetic gene cluster in Salinispora tropica, *Chembiochem* 14, 955-962.
13. Zhang, W., Watanabe, K., Wang, C. C., and Tang, Y. (2007) Investigation of early tailoring reactions in the oxytetracycline biosynthetic pathway, *J. Biol. Chem.* 282, 25717-25725.
14. Kalaitzis, J. A., Cheng, Q., Thomas, P. M., Kelleher, N. L., and Moore, B. S. (2009) In vitro biosynthesis of unnatural enterocin and wailupemycin polyketides, *J. Nat. Prod.* 72, 469-472.
15. Zhang, W., Ames, B. D., Tsai, S. C., and Tang, Y. (2006) Engineered biosynthesis of a novel amidated polyketide, using the malonamyl-specific initiation module

- from the oxytetracycline polyketide synthase, *Appl. Environ. Microbiol.* 72, 2573-2580.
16. Xu, Z., Metsa-Ketela, M., and Hertweck, C. (2009) Ketosynthase III as a gateway to engineering the biosynthesis of antitumoral benastatin derivatives, *J. Biotechnol.* 140, 107-113.
 17. Meurer, G., and Hutchinson, C. R. (1995) Functional analysis of putative beta-ketoacyl:acyl carrier protein synthase and acyltransferase active site motifs in a type II polyketide synthase of *Streptomyces glaucescens*, *J. Bacteriol.* 177, 477-481.
 18. Meurer, G., Gerlitz, M., Wendt-Pienkowski, E., Vining, L. C., Rohr, J., and Hutchinson, C. R. (1997) Iterative type II polyketide synthases, cyclases and ketoreductases exhibit context-dependent behavior in the biosynthesis of linear and angular decapolyketides, *Chem. Biol.* 4, 433-443.
 19. Otten, S. L., Stutzman-Engwall, K. J., and Hutchinson, C. R. (1990) Cloning and expression of daunorubicin biosynthesis genes from *Streptomyces peucetius* and *S. peucetius* subsp. *caesius*, *J. Bacteriol.* 172, 3427-3434.
 20. Ye, J., Dickens, M. L., Plater, R., Li, Y., Lawrence, J., and Strohl, W. R. (1994) Isolation and sequence analysis of polyketide synthase genes from the daunomycin-producing *Streptomyces* sp. strain C5, *J. Bacteriol.* 176, 6270-6280.
 21. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J., and Leadlay, P. F. (1990) An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*, *Nature* 348, 176-178.
 22. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J., and Katz, L. (1991) Modular organization of genes required for complex polyketide biosynthesis, *Science* 252, 675-679.
 23. Grimm, A., Madduri, K., Ali, A., and Hutchinson, C. R. (1994) Characterization of the *Streptomyces peucetius* ATCC 29050 genes encoding doxorubicin polyketide synthase, *Gene* 151, 1-10.
 24. Rajgarhia, V. B., Priestley, N. D., and Strohl, W. R. (2001) The product of *dpsC* confers starter unit fidelity upon the daunorubicin polyketide synthase of *Streptomyces* sp. strain C5, *Metab. Eng.* 3, 49-63.
 25. Bao, W., Sheldon, P. J., Wendt-Pienkowski, E., and Hutchinson, C. R. (1999) The *Streptomyces peucetius* *dpsC* gene determines the choice of starter unit in biosynthesis of the daunorubicin polyketide, *J. Bacteriol.* 181, 4690-4695.
 26. Bao, W., Sheldon, P. J., and Hutchinson, C. R. (1999) Purification and properties of the *Streptomyces peucetius* *DpsC* beta-ketoacyl:acyl carrier protein synthase III that specifies the propionate-starter unit for type II polyketide biosynthesis, *Biochemistry* 38, 9752-9757.
 27. Pan, H., Tsai, S., Meadows, E. S., Miercke, L. J., Keatinge-Clay, A. T., O'Connell, J., Khosla, C., and Stroud, R. M. (2002) Crystal structure of the priming beta-ketosynthase from the R1128 polyketide biosynthetic pathway, *Structure* 10, 1559-1568.
 28. Bretschneider, T., Zocher, G., Unger, M., Scherlach, K., Stehle, T., and Hertweck, C. (2012) A ketosynthase homolog uses malonyl units to form esters in cervimycin biosynthesis, *Nat. Chem. Biol.* 8, 154-161.

29. Molnos, J., Gardiner, R., Dale, G. E., and Lange, R. (2003) A continuous coupled enzyme assay for bacterial malonyl-CoA:acyl carrier protein transacylase (FabD), *Anal. Biochem.* 319, 171-176.
30. Zhang, Y. M., Rao, M. S., Heath, R. J., Price, A. C., Olson, A. J., Rock, C. O., and White, S. W. (2001) Identification and analysis of the acyl carrier protein (ACP) docking site on beta-ketoacyl-ACP synthase III, *J. Biol. Chem.* 276, 8231-8238.
31. Liu, W., Han, C., Hu, L., Chen, K., Shen, X., and Jiang, H. (2006) Characterization and inhibitor discovery of one novel malonyl-CoA: acyl carrier protein transacylase (MCAT) from *Helicobacter pylori*, *FEBS Lett.* 580, 697-702.
32. Beltran-Alvarez, P., Cox, R. J., Crosby, J., and Simpson, T. J. (2007) Dissecting the component reactions catalyzed by the actinorhodin minimal polyketide synthase, *Biochemistry* 46, 14672-14681.
33. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol.* 372, 774-797.
34. Sachdeva, S., Musayev, F., Alhamadsheh, M. M., Neel Scarsdale, J., Tonie Wright, H., and Reynolds, K. A. (2008) Probing reactivity and substrate specificity of both subunits of the dimeric *Mycobacterium tuberculosis* FabH using alkyl-CoA disulfide inhibitors and acyl-CoA substrates, *Bioorg. Chem.* 36, 85-90.
35. Kortemme, T., and Creighton, T. E. (1995) Ionisation of cysteine residues at the termini of model alpha-helical peptides. Relevance to unusual thiol pKa values in proteins of the thioredoxin family, *J. Mol. Biol.* 253, 799-812.
36. Hol, W. G., van Duijnen, P. T., and Berendsen, H. J. (1978) The alpha-helix dipole and the properties of proteins, *Nature* 273, 443-446.
37. Serre, L., Verbree, E. C., Dauter, Z., Stuitje, A. R., and Derewenda, Z. S. (1995) 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.
38. Polgar, L. (2005) The catalytic triad of serine peptidases, *Cell. Mol. Life Sci.* 62, 2161-2172.
39. Qiu, X., Janson, C. A., Konstantinidis, A. K., Nwagwu, S., Silverman, C., Smith, W. W., Khandekar, S., Lonsdale, J., and Abdel-Meguid, S. S. (1999) Crystal structure of beta-ketoacyl-acyl carrier protein synthase III. A key condensing enzyme in bacterial fatty acid biosynthesis, *J. Biol. Chem.* 274, 36465-36471.
40. Maloney, F. P., Gerwick, L., Gerwick, W. H., Sherman, D. H., and Smith, J. L. (2016) Anatomy of the beta-branching enzyme of polyketide biosynthesis and its interaction with an acyl-ACP substrate, *Proc. Natl. Acad. Sci. U. S. A.* 113, 10316-10321.
41. Miyanaga, A., Iwasawa, S., Shinohara, Y., Kudo, F., and Eguchi, T. (2016) Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis, *Proc. Natl. Acad. Sci. U. S. A.* 113, 1802-1807.
42. Crosby, J., Byrom, K. J., Hitchman, T. S., Cox, R. J., Crump, M. P., Findlow, I. S., Bibb, M. J., and Simpson, T. J. (1998) Acylation of *Streptomyces* type II polyketide synthase acyl carrier proteins, *FEBS Lett.* 433, 132-138.
43. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Meth. Enzymol.* 276, 307-326.

44. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J. Appl. Crystallogr.* **40**, 658-674.
45. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132.
46. Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard, *Acta Crystallogr. D Biol. Crystallogr.* **64**, 61-69.
47. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine, *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352-367.
48. Moriarty, N. W., Grosse-Kunstleve, R. W., and Adams, P. D. (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation, *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074-1080.
49. Meier, J. L., Mercer, A. C., Rivera, H., Jr., and Burkart, M. D. (2006) Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification, *J. Am. Chem. Soc.* **128**, 12174-12184.
50. Shakyia, G., Rivera, H., Jr., Lee, D. J., Jaremko, M. J., La Clair, J. J., Fox, D. T., Haushalter, R. W., Schaub, A. J., Bruegger, J., Barajas, J. F., White, A. R., Kaur, P., Gwozdzowski, E. R., Wong, F., Tsai, S. C., and Burkart, M. D. (2014) Modeling linear and cyclic PKS intermediates through atom replacement, *J. Am. Chem. Soc.* **136**, 16792-16799.
51. Soding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction, *Nucleic Acids Res.* **33**, W244-W248.

B

Acyltransferase (AT) activity

Ketosynthase (KS) activity

Figure 1. (A) Starter unit diversity in type II polyketides. Daunorubicin, oxytetracycline, benastatin A, enterocin, lomaiviticin A, and R1128a are all type II PKS products with novel starter units (shown in bold). (B) The role of DpsC in daunorubicin biosynthesis. DpsC uses the substrate propionyl-CoA to form an acyl-enzyme intermediate, which reacts with PPant thiol of *holo*-ACP (AT activity) or with malonyl-ACP (KS activity). Both enzymatic activities ultimately generate a 21-carbon poly- β -ketone intermediate, which is processed to yield daunorubicin.

Figure 2. The DpsC crystal structure and secondary elements. (A) Front view of the DpsC homodimer with monomer A in green and monomer B in blue. (B) Top view of the DpsC homodimer. (C) and (D) The secondary structural elements of DpsC. The DpsC dimer is shown with monomer A colored by secondary structure and monomer B outlined in black and white. α -Helices are in cyan, β -sheets are in red, and loop regions are in magenta. The double β - α - β - α - β - β core motifs are displayed on the left. The active site serine is shown in spheres and the secondary structure elements of the active site pocket entrance are displayed on the right.

Figure 3. The crystal structures and active sites of DpsC, FabH, ZhuH, and CerJ with monomers A and B in different shades of green and blue, respectively. (A) DpsC. (B) FabH. (C) ZhuH. (D) CerJ. (E) The DpsC, FabH, ZhuH, and CerJ active sites. The DpsC and CerJ active sites have the same overall orientation of the catalytic triad. In comparison, DpsC has serine and CerJ has cysteine as the active site nucleophile. The orientation of active site residues of ZhuH and FabH are different from DpsC and CerJ. ZhuH and FabH both contain cysteine as the active site nucleophile, but their active site histidine is not engaged in hydrogen bonding with the cysteine. In ZhuH and FabH, the histidine and asparagine are important for decarboxylation of malonyl-ACP, not acyl-transfer. (F) The DpsC active site contains a Ser-His-Asp catalytic triad that is responsible for acyl transfer. In the proposed mechanism of DpsC acyl transfer, Ser118 is activated as a nucleophile for acylation because of a neighboring His297 that forms a hydrogen bond with the hydroxyl proton on Ser118. The His297 is further stabilized by a neighboring aspartate.

Figure 4. The active sites architectures of *apo*-DpsC (A), acetyl-DpsC (B), propionyl-DpsC (C), and butyryl-DpsC (D). Entire active site architecture is shown in sticks (top) and space filling models in the same orientation (bottom) showing Leu93 acting as an active site cap to restrict the acyl substrate size. Distances are in Å.

Figure 5. The DpsC active site and residues that interact with the PPant analogue **4**. (A) Chemical structure of **4** (B) SA omit maps of **4** bound to DpsC (monomer A, left, and monomer B, right). (C) The substrate analogue **4** is bound in the active site of DpsC by a series of hydrogen bonds, hydrophobic interactions, and charge-charge interactions. Lys221 and Arg217 form salt bridges to the phosphate moiety of the PPant analogue at the solvent exposed face. The hydrophobic portions of the PPant analogue interact with Met197, Val234, Phe331, and Val327. The propionyl moiety sits in the active site with the amide carbonyl forming hydrogen bonds the amide hydrogens of the backbone residues Ala329 and Ala118. The hydrophobic portion of the propionyl moiety forms hydrophobic contacts in the acyl-binding region with Leu86, Leu93, Val117, and Leu217. (D) A view looking from the solvent towards the active site tunnel of DpsC. Lys221 (α -7) and Arg271 (α -10) anchor the phosphate of **4** and the dimethyl moiety of **4** sits between α -7 and α -8. (E) A series of water molecules form hydrogen bonds with the PPant moiety. (F) An overlay of the propionyl-DpsC and DpsC co-crystal with **4** showing similar orientations of the propionyl group.

B

Enzyme	Organism	Name of NP	Type of NP	% Identity to DpsC	Active Site Triad
DpsC	<i>Streptomyces peucetius</i>	daunorubicin	type II polyketide	100%	Ser-His-Asp
CosE	<i>Streptomyces olindensis</i>	cosmomycin	type II polyketide	61.6	Ser-His-Asp
CerJ	<i>Streptomyces tendae</i>	cervimycin	type II polyketide	41.3	Cys-His-Asp
SsfG	<i>Streptomyces SF2575</i>	SF2575	type II polyketide	32.7	Cys-His-Asp
PokM2	<i>Streptomyces diastatochromogenes</i>	polyketomycin	type II polyketide	33.6	Cys-His-Asp
PlaP2	<i>Streptomyces TU6071</i>	phenalinolactone	terpene	35.7	Cys-His-Asp
ChlB3	<i>Streptomyces antibioticus</i>	chlorothricin	type I polyketide	34.0	Cys-His-Asp
CalO4	<i>Micromonospora echinospora</i>	calicheamycin	type I polyketide	30.8	Cys-His-Asp
AviN	<i>Streptomyces viridochromogenes</i>	avilamycin	polysaccharide	34.7	Cys-Thr-Asp

Figure 6. Glycosylated natural products containing an acylated hydroxyl or thiol and putative acyltransferase enzymes identified from a protein BLAST search of DpsC. (A) Natural products with DpsC homologues present in their gene clusters. Possible acylation sites are highlighted in bold. (B) A table describing DpsC homologues identified by a protein BLAST from known gene clusters that produce products that contain acylated substructures and predictions of their active site catalytic triads based on sequence alignments and homology modelling.

Figure 7. Analysis of active site and acyl binding region residues of DpsC homologues. (A) A table comparing the active site, middle pocket, and active site cap residues of DpsC homologues. (B) The DpsC active site displaying residues in the acyl binding region. (C) The active site of the AviN homology model showing the effect of smaller middle pocket residues and active site cap on pocket size. The AviN active site is much larger than the DpsC active site and contains Phe211, which may form π - π stacking interaction with an aromatic substrate.

Supporting Information

Structural and Functional Studies of the Daunorubicin Priming Ketosynthase

DpsC

David R. Jackson¹, Gaurav Shakya¹, Avinash B. Patel¹, Lina Y. Mohammed², Kostas Vasilakis², Pakorn Wattana-Amorn², Timothy R. Valentic¹, Jacob Milligan¹, Matthew P. Crump², John Crosby², and Shiou-Chuan Tsai¹

1. Departments of Molecular Biology and Biochemistry, Chemistry, and Pharmaceutical Sciences, University of California, Irvine, Irvine, CA 92697.

2. School of Chemistry, University of Bristol, United Kingdom, Bristol BS8 1TS, United Kingdom.

Corresponding authors: Shiou-Chuan Tsai. Department of Molecular Biology and Biochemistry, Department of Chemistry, and Department of Pharmaceutical Sciences, University of California, Irvine. 949-824-4486. sctsai@uci.edu

John Crosby. School of Chemistry, University of Bristol, United Kingdom. +44 (0) 117 928 8445. John.Crosby@bristol.ac.uk

Matthew P. Crump. School of Chemistry, University of Bristol, United Kingdom. +44 (0) 117 331 7163. matt.crump@bristol.ac.uk

Table of Contents

Table S1. DpsC crystallographic data collection and refinement statistics.

Figure S1. DpsC self-acylation with different acyl-CoA substrates.

Figure S2. *In vitro* kinetic assay of DpsC self-acylation.

Figure S3. *In vitro* assays for DpsC acyl-transfer to DpsG.

Figure S4. The differences in spatial orientation between the DpsC and *EcFabH* active site residues.

Figure S5. SA omit maps of the DpsC active site.

Figure S6. Comparison of the DpsC active site with the *EcFabH* active site focusing on Leu93 (DpsC) and Phe87 (*EcFabH*).

Figure S7. Synthesis of the phosphorylated PPant analogue **4**.

Figure S8. Sequence alignment of putative DpsC homologues.

Figure S9. ^1H NMR assignment of **2**

Figure S10. ^{13}C NMR spectrum of **2**

Figure S11. ^1H NMR assignment of **3**

Figure S12. ^{13}C NMR spectrum of **3**

	<i>apo</i> -DpsC	acetyl-DpsC	propionyl-DpsC	butyryl-DpsC	DpsC co-crystal with 4
PDB ID:	5TT4	4XS7	4XSA	4XSB	4XS9
Crystallization	0.18 M sodium citrate, 26 % PEG 3350	0.18 M sodium citrate, 26 % PEG 3350	0.18 M sodium citrate, 26 % PEG 3350	0.18 M sodium citrate, 26 % PEG 3350	0.18 M sodium citrate, 26 % PEG 3350
Crystallographic Data					
Wavelength (Å)	0.9999	0.9775	1.0000	0.9999	0.9795
Space Group	P6 ₅ 22	P6 ₅ 22	P6 ₅ 22	P6 ₅ 22	P6 ₅ 22
Cell Dimension a, b, c (Å)	90.033, 90.033, 304.689	91.213, 91.213, 315.643	91.477, 91.477, 316.833	91.382, 91.382, 315.783	89.702, 89.702, 302.704
	$\alpha=\beta=90^\circ$, $\gamma=120^\circ$	$\alpha=\beta=90^\circ$, $\gamma=120^\circ$	$\alpha=\beta=90^\circ$, $\gamma=120^\circ$	$\alpha=\beta=90^\circ$, $\gamma=120^\circ$	$\alpha=\beta=90^\circ$, $\gamma=120^\circ$
Resolution (Å)	50.00 - 2.50	50.00-2.40	50.00 - 2.40	50.00 - 2.20	50.00-2.30
No. of observations	557086	500940	857439	411366	479239
No. of unique observations	26403	31516	40753	40709	33191
Completeness %	99.3 (98.8)	100.00 (100.00)	99.6 (99.2)	99.9 (100)	100.0 (100.0)
I/ σ (I)	24.1 (6.0)	18.6 (4.3)	25.0 (5.3)	18.4 (2.6)	19.8 (6.0)
R _{merge} %	14.2 (65.0)	13.9 (79.9)	16.5 (75.9)	11.3 (81.9)	17.8 (44.3)
Redundancy	21.1	15.9	22.5	10.1	14.4
Refinement					
Resolution (Å)	50.00 - 2.50 (2.59 - 2.50)	50.00 - 2.40 (2.49 - 2.40)	50.00 - 2.4 (2.49 - 2.40)	50.00 - 2.20 (2.28 - 2.20)	50.00 - 2.30 (2.38 - 2.30)
No. of protein atoms	5058	5068	5102	5104	5073
No. of ligand atoms	0	0	0	0	52
No. of water atoms	173	248	572	431	316
R _{free} %	20.88 (23.47)	21.57 (29.88)	20.29 (24.10)	22.49 (31.98)	19.76 (23.02)
R _{crys} %	16.18 (18.96)	17.70 (20.80)	16.64 (18.10)	17.79 (25.03)	15.93 (17.81)
Geometry					
RMS bonds (Å)	0.008	0.023	0.009	0.008	0.008
RMS angles (°)	1.09	1.40	1.17	1.16	1.10
Ramachandran Favored (%)	98.00	96.00	97.00	96.00	98.00
Ramachandran Allowed (%)	1.55	3.55	2.56	3.41	1.70
Ramachandran Disallowed (%)	0.45	0.45	0.44	0.59	0.30
Average B-factors (Å²)					
Protein	32.40	34.30	17.50	29.90	29.00
Water	33.80	35.30	26.40	35.70	40.20
Ligands	N/A	N/A	N/A	N/A	33.10

Table S1. DpsC crystallographic data collection and refinement statistics.

*Numbers in parentheses denote the highest resolution shell.

Figure S1. DpsC self-acylation with different acyl-CoA substrates. (A) Deconvoluted ESI-MS spectra of purified DpsC after incubation with different CoA derivatives (acetyl-CoA, propionyl-CoA, butyryl-CoA and malonyl-CoA). ESI-MS analysis indicated that DpsC does not selectively load propionate. Calculated masses of DpsC, acetyl-DpsC, propionyl-DpsC, butyryl-DpsC and malonyl-DpsC are 39248, 39287, 39304, 39315 and 39334 Da respectively. (B) Graph showing the self-acylation efficiency of DpsC towards different starter units in non-saturating conditions. Higher affinity for propionate was not observed.

B

Kinetic parameter	acetyl-CoA	propionyl-CoA	butyryl-CoA	malonyl-CoA
$V_{\max}$ (μM.min ⁻¹)	2.31 ± 0.06	2.91 ± 0.07	2.24 ± 0.03	1.15 ± 0.04
K_M (μM)	46.14 ± 4.81	50.21 ± 4.35	44.96 ± 2.72	11.35 ± 2.21
k_{cat} (min ⁻¹)	0.12	0.14	0.11	0.06
k_{cat}/K_M (μM ⁻¹ .min ⁻¹)	2.6x10 ⁻³ ± 3x10 ⁻⁴	2.8x10 ⁻³ ± 2x10 ⁻⁴	2.6x10 ⁻³ ± 2x10 ⁻⁴	5.3x10 ⁻³ ± 1x10 ⁻³

Figure S2. *In vitro* kinetic assay of DpsC self-acylation. (A) Schematic representation of the mechanism of KDH assay and Michaelis-Menten plots for the self-acylation of DpsC. (B) Kinetic parameters were calculated by fitting the Michaelis-Menten curve to the experimental data.

Figure S3. *In vitro* assays for DpsC acyl-transfer to DpsG. (A) Deconvolved ESI-MS spectrum showing acyltransferase activity of DpsC when incubated with *holo*-DpsG and CoA derivatives. Calculated masses for *holo*-, acetyl-, propionyl- and butyryl-DpsG are 9604, 9646, 9660 and 9674 Da respectively. (B) The rate of conversion of *holo*-DpsG to the acylated form for all starter units. (C) Deconvolved ESI-MS spectrum of *holo*-DpsG components from acyltransferase activity by DpsC. No selective transfer of propionate onto DpsG was observed when the enzymes were incubated in a pool of different starter units.

Figure S4. The differences in spatial orientation between the DpsC and *EcFabH* active site residues. (A) In DpsC, Ser118, His297, and Asp302, are located on L5, L12, and Ca3, respectively. (B) In *EcFabH*, Cys112, His244, and Asn274, are located on L5, L12, and L11, respectively. (C) His297 and Asp302 of DpsC are behind the active site pocket and His297 is positioned so that the aromatic ring stacks against the side of the pocket. (D) His244 and Asn274 of *EcFabH* are positioned to the side of the active site pocket with both of their side chains able to form hydrogen bonds with substrates in the active site.

Figure S5. SA omit maps of the DpsC active site. The active site of monomer A is on the left (residues labeled) and the active site of monomer B is on the right (residues unlabeled). Acyl groups are colored in cyan. (A) Apo-DpsC. (B) Acetyl-DpsC. (C) Propionyl-DpsC. (D) Butyryl-DpsC.

Figure S6. Comparison of the DpsC active site with the *EcFabH* active site focusing on Leu93 (DpsC) and Phe87 (*EcFabH*).

Figure S7. Synthesis of the phosphorylated PPant analogue **4**. Protected pantetheine amine **1** can be coupled with a propanoic acid, and deprotection yields the pantetheine analogue **3** with an amide linkage. **3** can be directly phosphorylated using the CoaA kinase to form a substrate mimic **4** for co-crystallization.

Figure S8. Sequence alignment of putative DpsC homologues. The active site residues are marked with a red dot and residues in the acyl binding region are marked with a blue dot.

Figure S9. ¹H NMR assignment of compound 2

Figure S10. ¹³C NMR spectrum of compound 2

Compound 2

Figure S11. ¹H NMR assignment of compound 3

Compound 3

Figure S12. ¹³C NMR spectrum of compound 3

Compound 3