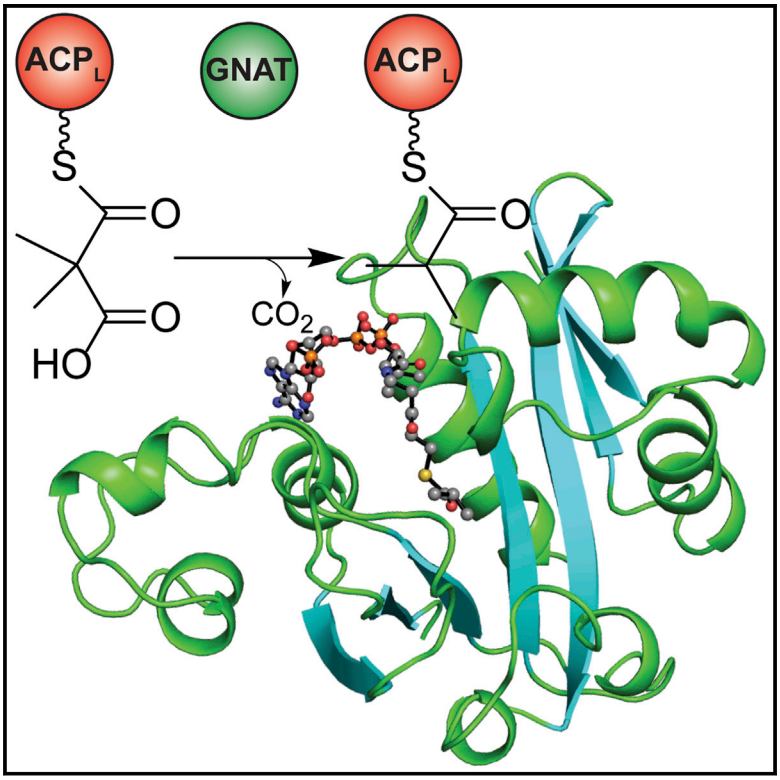
Article

Structure

Repurposing the GNAT Fold in the Initiation of Polyketide Biosynthesis

Graphical Abstract



Authors

Meredith A. Skiba, Collin L. Tran, Qingyun Dan, ..., William H. Gerwick, David H. Sherman, Janet L. Smith

Correspondence

janetsmith@umich.edu

In Brief

Most GCN5-related N-acetyltransferase (GNAT) superfamily members catalyze acyltransfer, but some are decarboxylases. Skiba et al. show that "GNATs" involved in the biosynthesis of two polyketide natural products catalyze only decarboxylation. They identify GNAT-like decarboxylases on distant nodes of the superfamily, illustrating the adaptability of the GNAT fold.

Highlights

- GNAT superfamily catalyzes decarboxylation in addition to acyltransfer
- A GNAT-like enzyme initiates biosynthesis of some polyketide natural products
- Polyketide synthase GNAT-like enzymes catalyze only decarboxylation

Skiba et al., 2020, Structure *28*, 1–12
January 7, 2020
https://doi.org/10.1016/j.str.2019.11.004

CellPress

Please cite this article in press as: Skiba et al., Repurposing the GNAT Fold in the Initiation of Polyketide Biosynthesis, Structure (2019), https://doi.org/10.1016/j.str.2019.11.004

Structure
Article

Repurposing the GNAT Fold in the Initiation of Polyketide Biosynthesis

Meredith A. Skiba,[1,2,8] Collin L. Tran,[1,9] Qingyun Dan,[1] Andrew P. Sikkema,[1,2,10] Zachary Klaver,[1] William H. Gerwick,[3,4] David H. Sherman,[1,5,6,7] and Janet L. Smith[1,2,11,*]

[1]Life Sciences Institute, University of Michigan, Ann Arbor, MI 48109, USA
[2]Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109, USA
[3]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093, USA
[4]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, USA
[5]Department of Chemistry, University of Michigan, Ann Arbor, MI 48109, USA
[6]Department of Medicinal Chemistry, University of Michigan, Ann Arbor, MI 48109, USA
[7]Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109, USA
[8]Present address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA
[9]Present address: Proteases and Tissue Remodeling Section, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD 20892, USA
[10]Present address: Epigenetics and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA
[11]Lead Contact
*Correspondence: janetsmith@umich.edu
https://doi.org/10.1016/j.str.2019.11.004

SUMMARY

Natural product biosynthetic pathways are replete with enzymes repurposed for new catalytic functions. In some modular polyketide synthase (PKS) pathways, a GCN5-related N-acetyltransferase (GNAT)-like enzyme with an additional decarboxylation function initiates biosynthesis. Here, we probe two PKS GNAT-like domains for the dual activities of *S*-acyl transfer from coenzyme A (CoA) to an acyl carrier protein (ACP) and decarboxylation. The GphF and CurA GNAT-like domains selectively decarboxylate substrates that yield the anticipated pathway starter units. The GphF enzyme lacks detectable acyl transfer activity, and a crystal structure with an isobutyryl-CoA product analog reveals a partially occluded acyltransfer acceptor site. Further analysis indicates that the CurA GNAT-like domain also catalyzes only decarboxylation, and the initial acyl transfer is catalyzed by an unidentified enzyme. Thus, PKS GNAT-like domains are re-classified as GNAT-like decarboxylases. Two other decarboxylases, malonyl-CoA decarboxylase and EryM, reside on distant nodes of the superfamily, illustrating the adaptability of the GNAT fold.

INTRODUCTION

Named for the founding member, general control non-repressible 5 (GCN5)-related N-acetyltransferase (GNAT) (Neuwald and Landsman, 1997), the GNAT superfamily spans all kingdoms of life and is best known for using an acetyl-coenzyme A (CoA) donor to acetylate diverse amine substrates, such as histones, aminoglycoside antibiotics, arylalkylamines, and a variety of other proteins and metabolic intermediates (Figure 1) (Couture and Trievel, 2006; Favrot et al., 2016; Salah Ud-Din et al., 2016). Other GNAT superfamily members use alternative acyl donors, such as myristoyl-CoA (Farazi et al., 2001), succinyl-CoA (Vetting et al., 2008), or fatty acyl-ACPs (Gould et al., 2004; Van Wagoner and Clardy, 2006; Watson et al., 2002).

Natural product biosynthetic pathways frequently repurpose enzymes found elsewhere in biology for the production of molecules that confer a selective advantage on the producing organism. For example, a GNAT-like enzyme was acquired in polyketide biosynthetic pathways of a variety of bacteria (Figures 2 and S3) (Chang et al., 2004; Grindberg et al., 2011; Gu et al., 2007; Kampa et al., 2013; Kellmann et al., 2008; Mattheus et al., 2010; Moebius et al., 2012; Partida-Martinez et al., 2007; Piel, 2002; Piel et al., 2004a; Simunovic et al., 2006; Young et al., 2013), and first identified in sequence analysis of the pederin biosynthetic pathway (Piel et al., 2004b). Pederin and other polyketides are synthesized in a stepwise manner from acyl-CoA building blocks. Intermediates in the modular type I polyketide synthase (PKS) enzyme assembly lines are tethered to ACP domains via the phosphopantetheine (Ppant) cofactor. PKS GNAT-like domains exist in "loading" or initiation modules and are expected to transfer an acetyl-group from acetyl-CoA to the terminal thiol acceptor of the Ppant cofactor of the initiation module ACP (ACP_L). Identification of a putative diphosphate-binding P loop for acetyl-CoA, which exists in other GNAT family members, suggested that the pederin GNAT-like domain shares the canonical acetyl-CoA binding property of the GNAT fold. Additional GNAT-like domains exist in PKS initiation modules that produce acetyl (Chang et al., 2004; Kampa et al., 2013; Mattheus et al., 2010; Moebius et al., 2012; Partida-Martinez et al., 2007; Piel et al., 2004a), propionyl (Kellmann et al., 2008; Simunovic et al., 2006), isobutyryl (Young

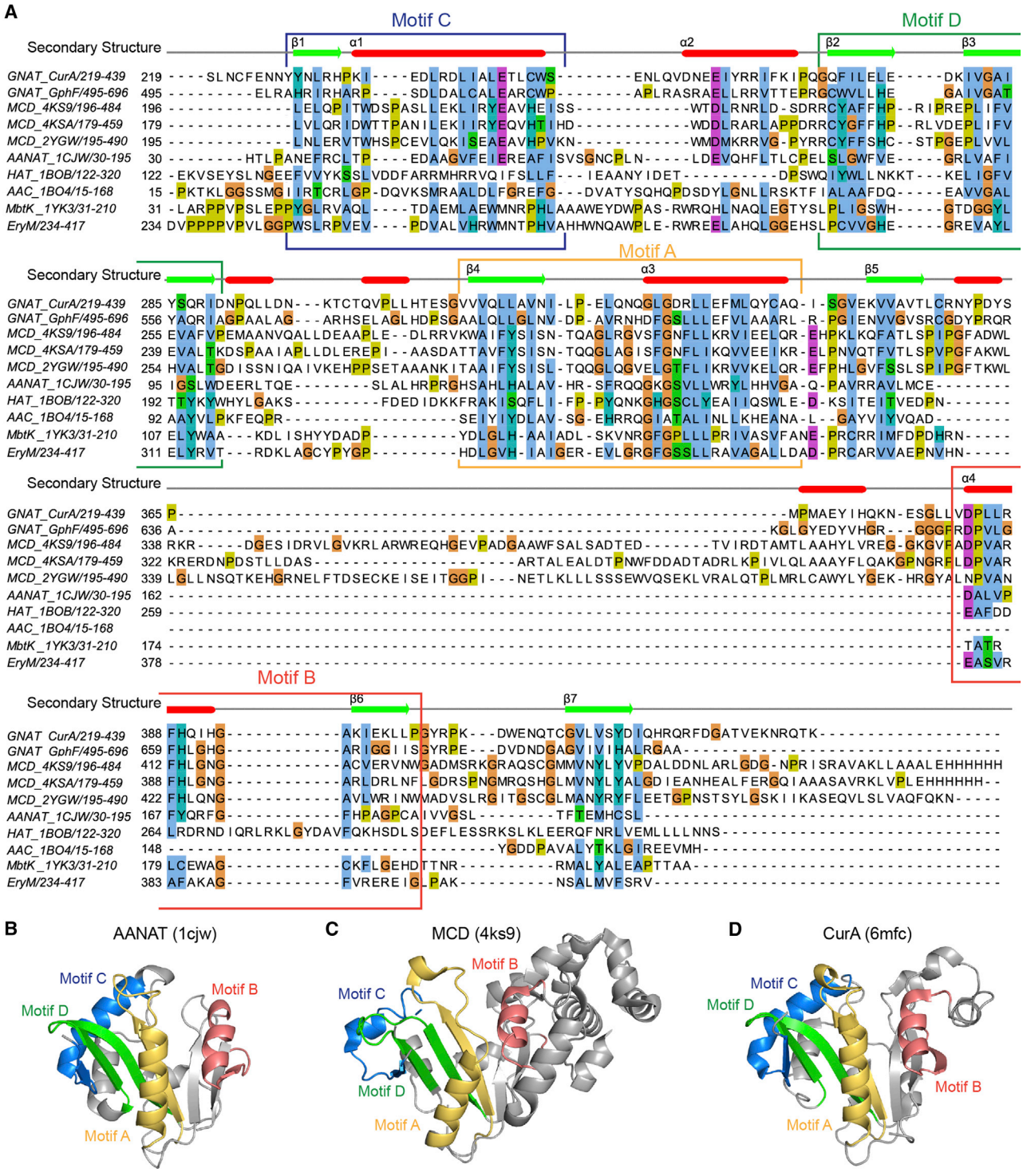


Figure 1. 3D Structures of GNAT Superfamily Members

(A) Structure-based sequence alignment of PKS GNAT-like domains, malonyl-CoA decarboxylase (MCD), and N-acetyltransferases (see also Figure S1). Structural motifs A–D of the GNAT fold are outlined in colored boxes (Neuwald and Landsman, 1997). The CurA GNAT secondary structure is displayed above the alignment. PKS GNAT-like domains and MCD contain an insertion within motif B. Protein abbreviations: Gph (gephyronic acid), PDB: 6MFC; Cur (curacin A), PDB: 2REE (Gu et al., 2007); *Cupriavidus metallidurans* MCD, PDB: 4KS9 (Froese et al., 2013); *Rhodopseudomonas palustris* MCD, PDB: 4KSA (Froese et al., 2013);

(legend continued on next page)

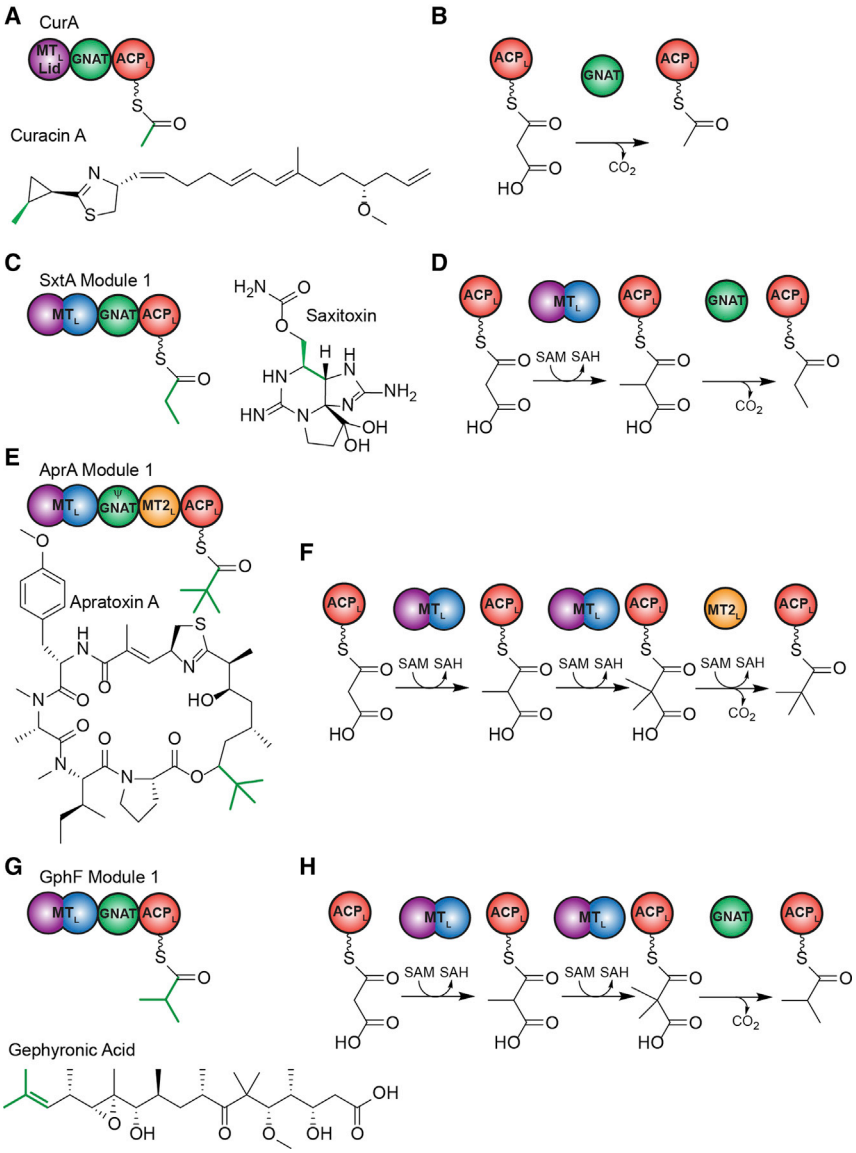


Figure 2. GNAT-like Enzymes in Curacin A, Saxitoxin, Apratoxin A, and Gephyronic Acid Biosynthesis

(A) CurA initiation module for an acetyl starter unit (green).

(B) CurA Mal-ACP decarboxylation.

(C) SxtA initiation module for a propionyl starter unit.

(D) SxtA Mal-ACP methylation and MeMal-ACP decarboxylation.

(E) AprA initiation module for a pivaloyl starter unit.

(F) AprA Mal-ACP dimethylation and coupled Me_2Mal-ACP decarboxylation/methylation.

(G) GphF initiation module for an isobutyryl starter unit.

(H) GphF Mal-ACP dimethylation and Me_2Mal-ACP decarboxylation (see also Figure S3).

et al., 2013), and pivaloyl (Grindberg et al., 2011; Skiba et al., 2018b; Sudek et al., 2007) starter units (Figures 2 and S3).

The initial characterization of a PKS GNAT-like domain was for the CurA GNAT in the pathway for the antimitotic metabolite curacin A, where we reported a slow *S*-acetyl transfer from acetyl-CoA to the CurA ACP_L to generate the acetyl starter unit of curacin A biosynthesis (Gu et al., 2007) (Figures 2A and 2B). Unexpectedly, CurA GNAT also rapidly decarboxylated malonyl (Mal)-ACP or Mal-CoA and thus could form acetyl-ACP directly from Mal-CoA. Decarboxylation represented a new catalytic function for the GNAT superfamily and required His and Thr amino acids (Gu et al., 2007). The apparent dual activities clarified the function of GNAT-like domains in the initiation modules of other PKS systems.

This was demonstrated in the saxitoxin pathway, where the SxtA GNAT-like domain (Figures 2C and 2D) also displayed dual activities, including a very weak acyl transfer activity and a decarboxylation activity with preference for methylmalonyl (MeMal)-ACP conversion to propionyl-ACP, the expected saxitoxin starter unit, over Mal-ACP (Chun et al., 2018). In contrast, in apratoxin A and bryostatin biosynthesis (Figures 2E and 2F), the GNAT-like domain is truncated and possesses no detectable decarboxylation or acyl transfer activity (Skiba et al., 2017, 2018b). In these pathways, a specialized methyltransferase domain catalyzes decarboxylation coupled to methyl transfer, and the initial acyl transfer step to commence apratoxin A biosynthesis requires an enzyme not encoded in the gene cluster such as FabD, the malonyl-acyltransferase of fatty acid biosynthesis (Skiba et al., 2018b).

Biosynthesis of the cytostatic polyketide gephyronic acid begins with the GphF initiation module (Young et al., 2013), which includes methyltransferase (MT), GNAT-like, and ACP domains (MT_L-$GNAT_L$-ACP_L, Figure 2G). An isobutyryl starter unit is the expected product of the GphF initiation module, in contrast to other characterized GNAT-containing initiation modules, which produce acetyl (curacin A) (Gu et al., 2007), propionyl (saxitoxin) (Chun et al., 2018), and pivaloyl (apratoxin and bryostatin) starter units (Skiba et al., 2018b). Isotope-labeling studies suggested that the gephyronic acid isobutyryl unit is derived from *S*-adenosylmethionine (SAM) (Young et al., 2013). Previously, we showed that GphF MT_L and other MTs associated with PKS GNAT-like domains are unusual mononuclear iron-dependent enzymes that catalyze methylation of Mal-ACP to

Homo sapiens MCD, PDB: 2YGW (Froese et al., 2013); serotonin N-acetyltransferase AANT, PDB: 1CJW (Hickman et al., 1999); histone acetyltransferase HAT, PDB: 1BOB (Dutnall et al., 1998); aminoglycoside 3-N-acetyltransferase AAC, PDB: 1BO4 (Wolf et al., 1998); *M. tuberculosis* N-acetyltransferase MbtK, PDB: 1YK3 (Card et al., 2005), *Saccharopolyspora erythraea* N-acetyltransferase/decarboxylase EryM (see also Figure S2.). Structures of GNAT superfamily members (B) serotonin N-acetyltransferase AANT, (C) *C. metallidurans* MCD, and (D) CurA.

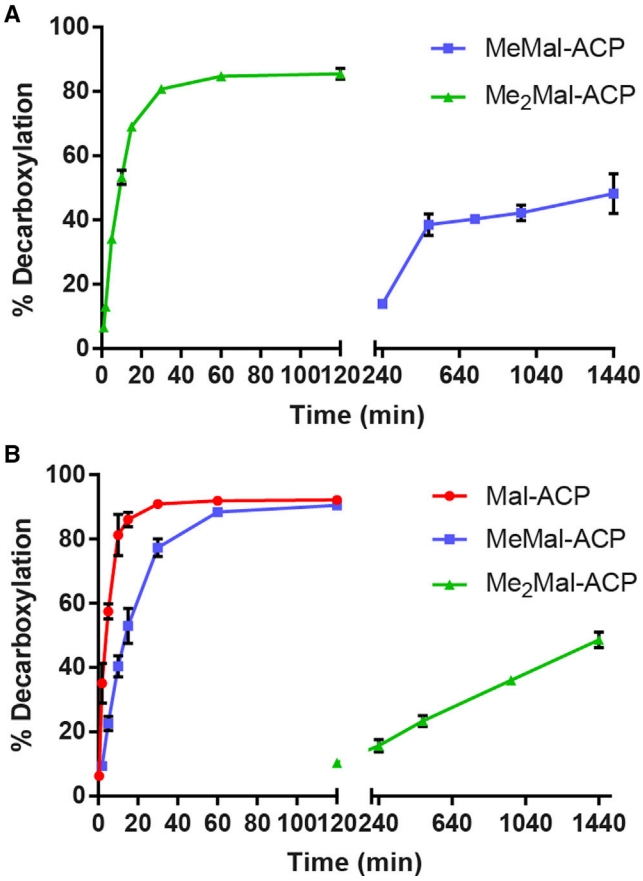


Figure 3. GphF and CurA GNAT Decarboxylation Activity

Decarboxylation of acyl-ACP substrates by (A) GphF GNAT and (B) CurA GNAT. Activity was monitored via the Ppant ejection assay (Dorrestein et al., 2006; Meluzzi et al., 2008). Error bars represent triplicate experiments and, in some cases, are too small to be visible. GphF GNAT did not decarboxylate Mal-ACP in a 48-h reaction time (see also Figure S4).

MeMal-ACP or dimethylmalonyl (Me_2Mal)-ACP (Figure 2H) (Skiba et al., 2017). We determined that the carboxylate of Mal-ACP is essential for methyl transfer activity, calling into question the presumed coupling of the acyl transfer and decarboxylation activities of GNAT-like domains. Decarboxylation of Me_2Mal-ACP by the GphF GNAT-like enzyme would result in the predicted isobutyryl starter unit, but the reaction cannot be coupled to malonyl transfer from CoA to ACP (Figure 2H).

Here, we investigate the catalytic function of two GNAT-like domains in polyketide biosynthesis. The GphF and CurA GNAT-like enzymes displayed a strong acyl group selectivity for their respective decarboxylation substrates. Additional insight into the decarboxylation reaction is provided by a 2.8-Å crystal structure of the GphF GNAT-like domain with the product analog isobutyryl-CoA. No acyl transfer activity was detected for the GphF GNAT-like domain. Given the low rate of acyl transfer compared with decarboxylation for PKS GNAT-like domains (Chun et al., 2018; Gu et al., 2007), their biological function appears to be primarily or exclusively decarboxylation. Thus, the GNAT-like domains in PKS pathways are more appropriately classified as GNAT-like decarboxylases.

RESULTS

GNAT-like Enzymes Encode Strict Substrate Specificity

We first investigated the decarboxylation activity of the GphF GNAT-like domain. Metal-dependent MT_L domains in PKS initiation modules can produce either a monomethylated (MeMal-ACP) (Chun et al., 2018) or dimethylated (Me_2Mal-ACP) (Skiba et al., 2017) product, which is presumably decarboxylated by the GNAT domain to produce the polyketide starter unit (Figure 2). Thus, the GNAT-like domain may be a gatekeeper, selectively decarboxylating the MT_L product with the pathway-specific number of methyl groups. The structure and pathway annotation of gephyronic acid (Young et al., 2013) as well as initial biochemical characterization (Skiba et al., 2017) indicated that the GphF MT_L converts Mal-ACP to Me_2Mal-ACP via a MeMal-ACP intermediate (Figure 2H). Thus, we tested the GphF GNAT selectivity for Me_2Mal-ACP, MeMal-ACP and Mal-ACP.

Mal-ACP and MeMal-ACP were generated from the respective CoAs via the promiscuous 4′-phosphopantetheinyl transferase Svp (Sánchez et al., 2001). Me_2Mal-ACP was generated enzymatically from MeMal-ACP using AprA MT_L from the apratoxin A initiation module (Skiba et al., 2017). The AprA ACP_L (Grindberg et al., 2011) was used as a surrogate for GphF ACP_L, which was insoluble. Relative rates of decarboxylation of the ACP-linked substrates were measured using the mass-spectrometry-based Ppant ejection assay (Dorrestein et al., 2006; Meluzzi et al., 2008). GphF GNAT displayed a strong selectivity for decarboxylation of Me_2Mal-ACP relative to MeMal-ACP (100-fold) or Mal-ACP (barely detectable within 48 h) (Figures 3A and S4). Thus, GphF GNAT requires the action of MT_L before decarboxylation and is selective for the Me_2Mal-ACP product over the MeMal-ACP intermediate, thereby priming the gephyronic acid assembly line with the isobutyryl-ACP starter group (Figure 2H).

We were curious to examine whether other PKS GNAT-like domains exhibit a similar degree of substrate selectivity. The CurA GNAT (Gu et al., 2007) is of particular interest as the CurA initiation module includes the vestige of an MT_L (N-terminal lid region formerly known as an adaptor) but lacks the SAM-binding core (Figure 2A), suggesting that CurA evolved from a module containing an MT_L (Skiba et al., 2017). CurA GNAT displayed a modest ~5-fold selectivity for its natural Mal-ACP substrate over MeMal-ACP (Figure 3B), in agreement with our previous report (Gu et al., 2007), and a 500-fold selectivity for Mal-ACP over Me_2Mal-ACP (Figure 3B). Although the selectivity for Mal-ACP over MeMal-ACP is weak, it is unlikely to result in mispriming with a propionyl-ACP starter unit, because many bacteria, including the *Moorea producens* cyanobacterial producer of curacin A, lack a gene for propionyl-CoA carboxylase (Lombard and Moreira, 2011) and do not produce MeMal-CoA at levels sufficient for secondary metabolite biosynthesis.

GphF GNAT Structure and Active Site

To understand the substrate selectivity of PKS GNAT-like domains for decarboxylation, we solved crystal structures of the GphF GNAT free enzyme (2.6 Å) and a binary complex with an isobutyryl-CoA product mimic (2.8 Å, Figure 4; Table 1). GphF

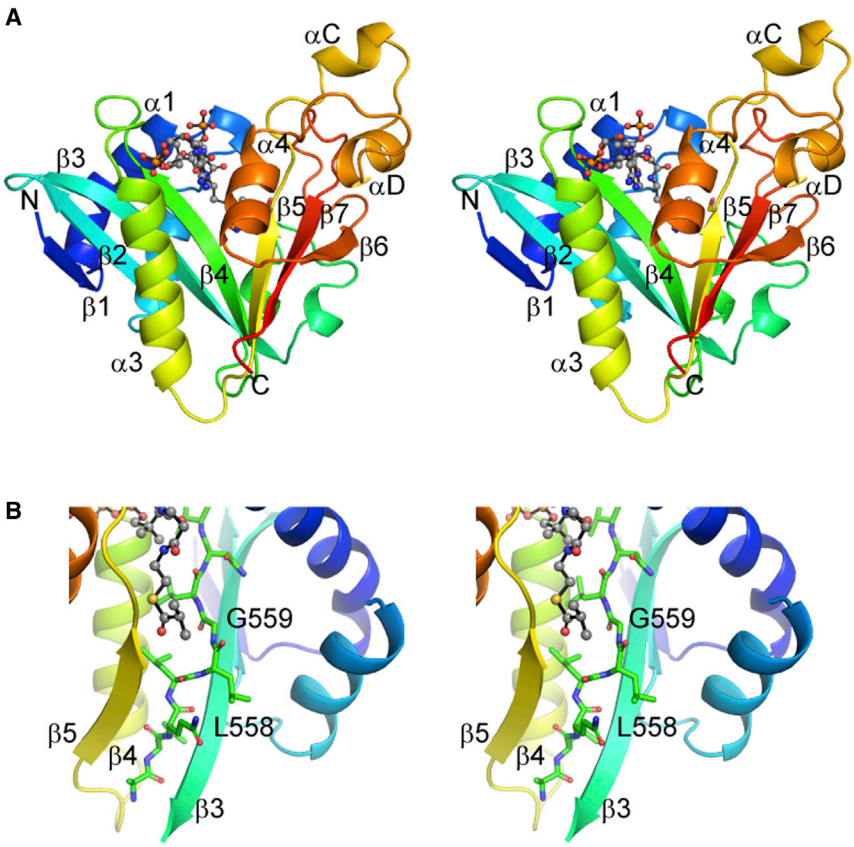


Figure 4. GphF GNAT Structure

(A) GphF GNAT is colored as a rainbow from blue (N terminus) to red (C terminus), shown in stereo. Isobutyryl-CoA (ball-and-stick form with atomic colors: gray C, red O, blue N, yellow S) binds to a V-shaped cleft between β4 (green) and β5 (yellow).

(B) Characteristic GNAT β-bulge in β4, rendered as sticks in this stereo image with coloring as in (A). The β-bulge is at the base of the V-shaped cleft between β4 (green) and β5, and places the α-carbons of Leu558 and Gly559 on the same face of the β sheet. The view is from the opposite face of the β sheet as in (A) (see also Figure S5).

GNAT displays the conserved GNAT fold comprised of a central β sheet flanked by α helices with a V-shaped Ppant binding cleft between parallel strands β4 and β5 (Figure 4A) (Vetting et al., 2005). A characteristic "bulge" in β4 (GphF Leu588) disrupts the hydrogen bonding scheme of the central β sheet, thereby creating the active site cleft (Figure 4B). The Ppant site includes a characteristic phosphate-binding loop (GphF 597–602 between β4 and α3, Figure 5), which is part of conserved "motif A" of the GNAT superfamily (Figure 1) (Neuwald and Landsman, 1997). Relative to GNAT N-acetyltransferases, PKS GNAT-like domains are distinguished by a helical insertion in GNAT motif B, following β5 (αC, αD) (Figures 4A and 5).

At the substrate entrance, the isobutyryl-CoA Ppant threads along the open edge of strand β4 where it splays away from β5 (Figure 4A). The thioester is positioned at the base of the V-shaped cleft near the conserved Thr/Ser and His residues (GphF Ser626, His660; CurA Thr355, His 389), which are essential for decarboxylation by CurA GNAT (Figure 6) (Gu et al., 2007). The catalytic His of GphF and CurA GNAT resides on helix α4 at the same topological location as a Tyr that is essential for acetyl transfer in serotonin N-acetyltransferase (Hickman et al., 1999). A distal binding pocket for acceptor substrates in canonical GNAT acyltransferases lies beyond the V-shaped cleft on the opposite side of the β sheet from the catalytic His and Thr/Ser. In PKS GNAT-like domains, the pocket includes conserved Trp and Arg side chains (GphF Trp520, Arg675; CurA Trp 249, Arg404). However, access for acceptor substrates through the distal end of the GphF pocket is restricted by His571 (Figure S5A). The analogous residue in CurA GNAT (Cys300) is less bulky, and the distal pocket would be accessible to a Ppant substrate (Figure S5B). In some PKS GNAT-like enzymes, even larger aromatic side chains at this position would seemingly block access for acceptor substrates (Figure 5).

GNAT Catalysis

The function of active site residues and the basis for the observed substrate selectivity of GphF and CurA GNAT domains (Figure 3) was probed via site-directed mutagenesis using Me_2Mal-, MeMal-, and Mal-ACP substrates (Figure 6). Like the CurA GNAT (Gu et al., 2007), the conserved active site His (GphF His660) and Ser/Thr (GphF Ser626) are essential for decarboxylation of both Me_2Mal-ACP and MeMal-ACP by GphF GNAT (Figures 5, 6A, and 6B). In contrast to the high levels of activity with Me_2Mal-ACP, no decarboxylation of Mal-ACP was detected in a 24-h incubation for wild-type GphF GNAT or any variant. Substitutions to the conserved Arg (GphF R675E and R675K) in the acyl pocket resulted in substantially reduced product formation (Figures 6A and 6B). Analogous substitutions in CurA Arg404 were similarly deleterious (Figures 6C and 6D), validating a role for the conserved Arg in decarboxylation. We conclude that the positively charged Arg interacts with the carboxylate of Me_2Mal- or Mal-ACP, given the order-of-magnitude reduction in product formation for the charge-reversal substitutions (GphF R675E and CurA R404E) and the substantial decrease for the more conservative substitutions (GphF R675K and CurA R404K). The Arg675 guanidinium is well positioned to interact with the carboxylate of the Me_2Mal-ACP substrate, based on the binary complex of the GphF GNAT with the isobutyryl-CoA product mimic (Figure S5C). The essential active site His and Thr/Ser are optimal candidates to stabilize an enolate intermediate formed during decarboxylation and to re-protonate the carbanion upon collapse of the enolate, as previously proposed (Figure S5D) (Gu et al., 2007).

We next examined amino acids in the Ppant binding cleft and acyl group pocket that differ in GphF and CurA GNAT to identify the basis for the differing substrate selectivity. The essential Thr side chain in the decarboxylation site of PKS GNAT-like enzymes (CurA Thr355) is more common than Ser

Table 1. Crystallographic Information

Protein	GphF GNAT	GphF GNAT
Ligand		isobutyryl-CoA
Data Collection		
Space group	$C222_1$	$C222_1$
Unit cell		
a, b, c (Å)	138.3, 145.7, 78.0	134.4, 145.5, 77.2
X-ray source	APS 23ID-B	APS 23ID-B
Wavelength (Å)	1.033	1.033
d_{min} (Å)	2.59 (2.68–2.59)[a]	2.79 (2.89–2.79)
R-merge	0.1847 (3.16)	0.1762 (1.97)
Inner shell R-merge	0.047 (7.67 Å)	0.035 (8.28 Å)
Avg I/σ(I)	11.7 (0.9)	9.87 (1.10)
Completeness (%)	99.6 (97.3)	99.3 (95.7)
Multiplicity	13.1 (13.4)	6.8 (6.9)
Total observations	327,190 (31,990)	128,891 (12,396)
Wilson B factor (Å^2)	76.9	69.2
$CC_{1/2}$	0.998 (0.463)	0.997 (0.644)
CC*	1.00 (0.795)	0.999 (0.885)
Refinement		
Data range (Å)	42.19–2.59	42.8–2.79
Reflections	24,860	19,043
R_{work}/R_{free} (%)	23.0/26.8	23.2/27.5
No. of non-hydrogen atoms	3,011	3,070
Protein	2,934	2,934
Ligands	35	122
Water	42	14
Amino acid residues	390	392
Deviation from ideality		
Bond lengths (Å)	0.004	0.002
Bond angles (°)	0.98	0.42
Average B factor (Å^2)	91.6	103.5
Protein	91.5	103.4
Ligands	112.8	111.6
Solvent	76.4	66.7
Ramachandran plot (%)		
Favored	98.2	96.9
Allowed	1.8	3.1
Outliers	0	0

[a]Values in parentheses designate the outer shell of data for each category.

(GphF Ser626) (Figures 5, 6B, and 6D). Among GNAT-like domains in pathways where the natural product is known, only the GphF initiation module produces an isobutyryl starter group. Reasoning that the smaller active site Ser may provide more space to accommodate dimethylmalonyl than the bulkier Thr, we produced GphF S626T and CurA T355S, but these substitutions had only modest effects on decarboxylation of all substrates and did not influence the selectivity for Me_2Mal-ACP by the GphF GNAT or Mal-ACP by the CurA GNAT (Figures 6A and 6C). We made several additional reciprocal substitutions at amino acid positions in the active site and substrate pocket in either or both enzymes (GphF H690Y, CurA Y419H; GphF R597Q, CurA Q326R; GphF R627L, CurA L356R; GphF H571C). All but GphF R627L, which was insoluble, yielded stable protein, but none altered the substrate selectivity of either enzyme (Figure 6).

GphF GNAT Does Not Catalyze Acyl Transfer

The distal acyl-binding pocket of the CurA GNAT extends from the V-shaped cleft to the surface of the protein and was proposed to accommodate the Ppant of the ACP acyl acceptor in CurA GNAT (Gu et al., 2007). However, access through the distal end of the GphF pocket is partially blocked by His571 (Figures S5A and S5C). We also noted that amino acids in the distal pocket are among the least conserved in PKS GNAT-like domains, which is not consistent with a common Ppant pocket for ACP acceptors. Together with the low levels of malonyl or acetyl transfer for previously characterized PKS GNAT-like domains in CurA (Gu et al., 2007) and SxtA (Chun et al., 2018), this led us to investigate the acyltransferase activity of the GphF and CurA GNATs.

The proposed reaction scheme for the GphF initiation module requires the transfer of malonate from CoA to ACP before the methylation reactions by MT_L. This ensures that the resulting Me_2Mal product is non-diffusable and available for further processing by subsequent enzymes in the gephyronic acid pathway. GphF GNAT is a candidate to provide the malonyl transfer step as it was unable to decarboxylate malonyl-ACP before methylation by GphF MT_L (Figure 2H). However, we detected extremely low levels of malonyl transfer activity by GphF GNAT, unlike the consistently high levels of Me_2Mal-ACP decarboxylation.

For both GphF GNAT and MT_L-GNAT, the consistently low levels of acyl transfer activity varied among apparently identical preparations, suggesting that the observed activity may be due to a contaminating enzyme from the *E. coli* expression host. Previously, contaminating components of the highly active *E. coli* fatty acid biosynthetic machinery were shown to transfer malonyl from CoA to PKS ACPs (Florova et al., 2002). To separate GphF GNAT (pI 7.3) from the *E. coli* fatty acid synthase enzymes most likely to support acyl transfer (FabD [pI 5.0] and FabH [pI 5.1]), we added an additional ion-exchange chromatography step to the purification protocol. The stringently purified GphF GNAT retained a high rate of Me_2Mal-ACP decarboxylation but did not support acyl transfer to ACP from a panel of acyl-CoA donors, including Mal-CoA and isobutyryl-CoA (Figures 7A and S6). Thus, we conclude that the GphF GNAT-like enzyme is not an acyltransferase.

The lack of acyl transfer activity for GphF GNAT prompted us to reinvestigate CurA GNAT acyl transfer, which was nearly 800-fold slower than decarboxylation (Gu et al., 2007). We performed additional affinity chromatography and size-exclusion purification steps and found that the multistep-purified CurA GNAT had 5-fold diminished acyl transfer activity compared with the single-step purified enzyme used previously (Gu et al., 2007) (Figures 7B and S7). The multistep-purified enzyme retained rapid decarboxylation activity (Figure S8).

To determine whether contaminating enzymes from the *E. coli* heterologous host could contribute to the observed acyltransferase activity for the PKS GNAT-like domains, we carried out

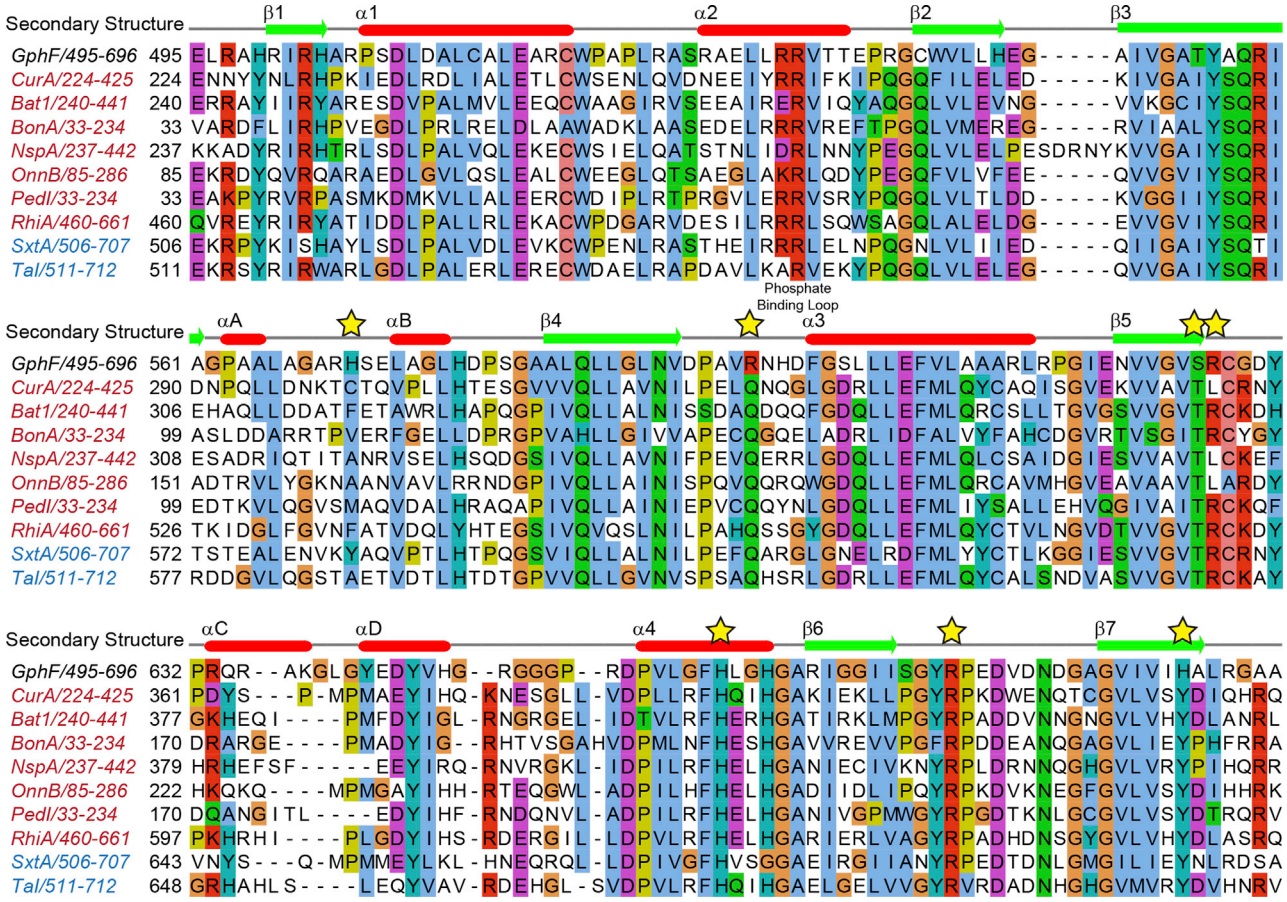


Figure 5. Sequence Alignment of GNAT-like Domains from PKS Pathways for Natural Products of Known Structure

All GNAT-like domains occur in presumed initiation modules. Only GphF is predicted to produce an isobutyryl starter unit; initiation modules with propionyl starter units are in blue and those with acetyl starter units in red. The GphF GNAT secondary structure annotation is displayed above the alignment. Amino acids subjected to mutagenesis are starred. Protein abbreviations (GenBank accession codes): Gph (gephyronic acid), KF479198.1; Cur (curacin A), AEE88289.1; Bat (batumin), WP_052451043.1; Bon (bongkrekic acid), AFN27480.1; Nsp (nosperin), ADA69237.1; Onn (onnamide), AAV97870.1; Ped (pederin), AAR19304.1; Rhi (rhizoxin), WP_013435483.1; Sxt (saxitoxin), WP_009343302.1; Ta (myxovirescin A), WP_011553948.1.

a mock purification of *E. coli* proteins from cells bearing an empty expression vector. The cytosolic extract enriched by the Ni-NTA affinity step was concentrated and subjected to acyl transfer assays with the CurA ACP (Figure S8). The enriched *E. coli* protein mixture catalyzed transfer of both acetyl and malonyl from CoA to CurA ACP in a concentration-dependent manner at rates faster than the multistep purified CurA GNAT, but the mixture of *E. coli* proteins had no detectable decarboxylation of Mal-ACP (Figures 8A and S8). The malonyl and acetyl transfer activity of *E. coli* proteins is evidence that contaminants in preparations from single-step Ni-NTA purification were responsible for both acyltransfer reactions.

As PKS GNAT-like domains possess only decarboxylation activity, an additional enzyme must catalyze the initial malonyl transfer step to initiate both gephyronic acid and curacin biosynthesis in the producing organisms. FabD is an attractive candidate to provide this step for PKS initiation modules containing a GNAT-like decarboxylation domain, as shown in the apratoxin A biosynthetic pathway (Skiba et al., 2018b). We tested this possibility with the *Moorea bouillonii* FabD, which is 97% identical to FabD from the curacin A producer *M. producens*. Recombinant *M. bouillonii* FabD was ~40,000-fold more effective in malonyl transfer from CoA to CurA ACP than was the CurA GNAT (65% transfer in 20 min with 25 nM FabD compared with 8% transfer in 4 h with 10 μM GNAT, Figures 7B, 8B, and S7).

DISCUSSION

PKS GNAT-like Domains Are Decarboxylases

This study demonstrates that the GNAT-like domain in the PKS loading module for gephyronic acid biosynthesis is a highly selective Me_2Mal-ACP decarboxylase and that it lacks acyl transfer activity. Based on the results for the myxobacterial GphF $GNAT_L$ and the cyanobacterial CurA $GNAT_L$, decarboxylation is the primary role of GNAT-like enzymes in PKS initiation modules. The rate of acyl transfer is non-existent or suspiciously low for these GNAT-like domains compared with their decarboxylation rates (Gu et al., 2007) and can be reduced further by additional purification steps (Figure 7B), Furthermore, the previously proposed binding site for the Ppant of the CurA ACP acyl acceptor (Gu et al., 2007) is partially occluded in GphF and

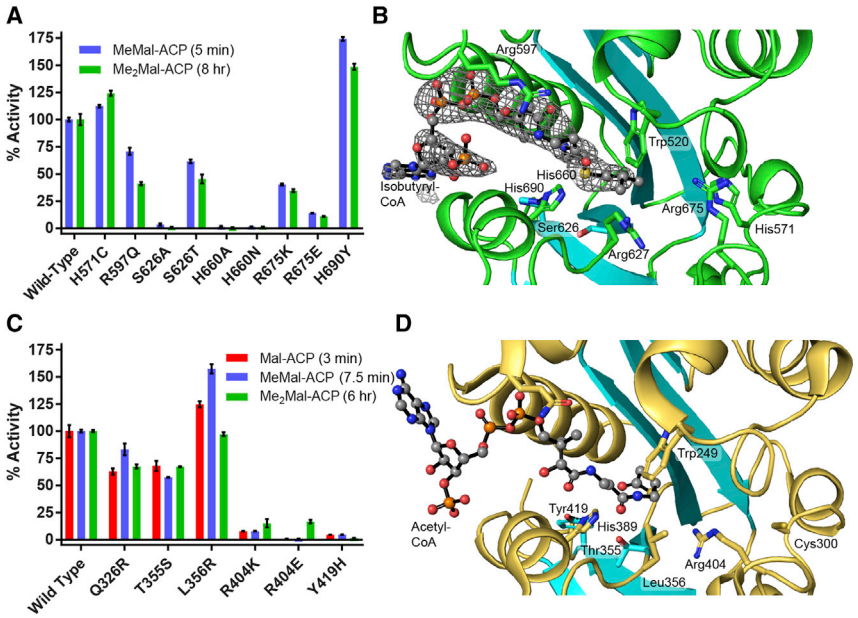


Figure 6. Probing Decarboxylation via Site-Directed Mutagenesis

(A) GphF GNAT decarboxylation reactions. Activity is normalized to wild-type levels at 5 min (Me_2Mal-ACP) and 8 h (MeMal-ACP) (Figure 3).

(B) GphF GNAT active site with bound isobutyryl-CoA. $F_o - F_c$ omit density for isobutyryl-CoA is shown in gray mesh at 3σ contour. Ser626 and His660 are the catalytic amino acids.

(C) CurA GNAT decarboxylation reactions. Activity is normalized to wild-type levels at 6 h (Me_2Mal-ACP), 7.5 min (MeMal-ACP), and 3 min (Mal-ACP) (Figure 3).

(D) CurA GNAT active site with bound acetyl-CoA (Gu et al., 2007). Thr355 and His389 are the catalytic amino acids. Error bars represent triplicate experiments and, in some cases, are too small to be observed. Bound CoAs are shown in ball-and-stick form. Amino acids subjected to mutagenesis are shown in sticks with atomic coloring (see also Figure S5).

may be fully occluded in other PKS GNAT-like domains (Figures 5 and S5). Thus, the PKS GNAT-like domain is more appropriately designated as a GNAT-like decarboxylase. Decarboxylation by these enzymes requires conserved His and Thr/Ser residues. We further demonstrated the importance of an active site Arg in decarboxylation by both cyanobacterial and myxobacterial GNAT-like decarboxylases; this amino acid is conserved among GNAT-like decarboxylases from PKS initiation modules with known biosynthetic products (Figures 5 and 6).

Our results suggest that the extremely slow and variable malonyl transfer activity of the GNAT-like decarboxylase or the MT_L-GNAT didomain is due to a contaminant from the *E. coli* expression host, as additional purification steps of the GNAT-like decarboxylases decreased acyl transfer activity but did not alter the rate of decarboxylation (Figures 7 and S8). Previously, the detection of an *E. coli* contaminant refuted a "self-loading" ability of PKS ACPs (Florova et al., 2002). We found that unidentified *E. coli* cytosolic proteins enriched by Ni-NTA resin support both malonyl and acetyl transfer from CoA to a PKS ACP. An obvious malonyl transferase candidate is *E. coli* FabD of fatty acid biosynthesis, and acetyl transfer activity has been detected in *E. coli* FabH (Tsay et al., 1992).

In PKS initiation modules containing GNAT-like decarboxylases, we propose that the ACP_L is primed with malonyl by a standalone malonyl-acyltransferase that may not be encoded in the PKS gene cluster. FabD, the malonyl-acyltransferase from fatty acid biosynthesis, is an obvious candidate, as it supported rapid malonyl acyl transfer to AprA ACP_L (Skiba et al., 2018b) and CurA ACP_L (Figure 8B), as well as other PKS ACPs (Florova et al., 2002; Kumar et al., 2003). Some GNAT-like domains exist in so-called *trans*-AT PKS pathways, where a standalone acyltransferase (AT) provides building blocks at all extension steps in the pathway. A *trans*-AT could catalyze the initiating acyl transfer step in such pathways (Kampa et al., 2013; Mattheus et al., 2010; Moebius et al., 2012; Partida-Martinez et al., 2007; Piel, 2002; Piel et al., 2004a; Simunovic et al., 2006).

Starter unit biosynthesis for several polyketide natural products is rationalized by the understanding of the decarboxylase function of GNAT-like PKS domains and by sequence analysis of the loading modules (Figure S3). After malonyl acyl transfer to the ACP domain, initiation modules that contain no MT_L, such as CurA (Figure 2A) and those of the pederin (Piel, 2002), onnamide (Piel et al., 2004a), and bongkrekic acid (Moebius et al., 2012) pathways, or those that contain a defunct and apparently vestigial MT_L, such as the rhizoxin (Partida-Martinez et al., 2007), batumin (Mattheus et al., 2010), and nosperin (Kampa et al., 2013) pathways, proceed directly to decarboxylation by the GNAT-like decarboxylase, yielding an acetyl starter unit. Initiation modules with an active MT_L domain catalyze a pathway-specific number of methyl transfers before decarboxylation by the GNAT-like domain, yielding an isobutyryl starter unit for gephyronic acid (Young et al., 2013) and a propionyl for saxitoxin (Kellmann et al., 2008) and myxovirescin (Simunovic et al., 2006). In yet another twist of nature, the GNAT-like variant in the apratoxin (Grindberg et al., 2011) and bryostatin (Sudek et al., 2007) pathways is inactive, and a second MT domain performs coupled methylation and decarboxylation, yielding a pivaloyl starter unit (Skiba et al., 2018b). In the pathways with active methyltransferase domains, methylation of Mal-ACP rather than Mal-CoA is essential, as MeMal and Me_2Mal are precious resources that must be sequestered by the PKS enzyme assembly line. The high degree of substrate selectivity of the GNAT-like decarboxylase limits the production of an incorrect starter unit, which may stall at subsequent biosynthetic steps or lead to an incorrect product. The crystal structures and mutagenesis studies of the GphF and CurA GNAT-like decarboxylases did not reveal the basis for the observed acyl group selectivity.

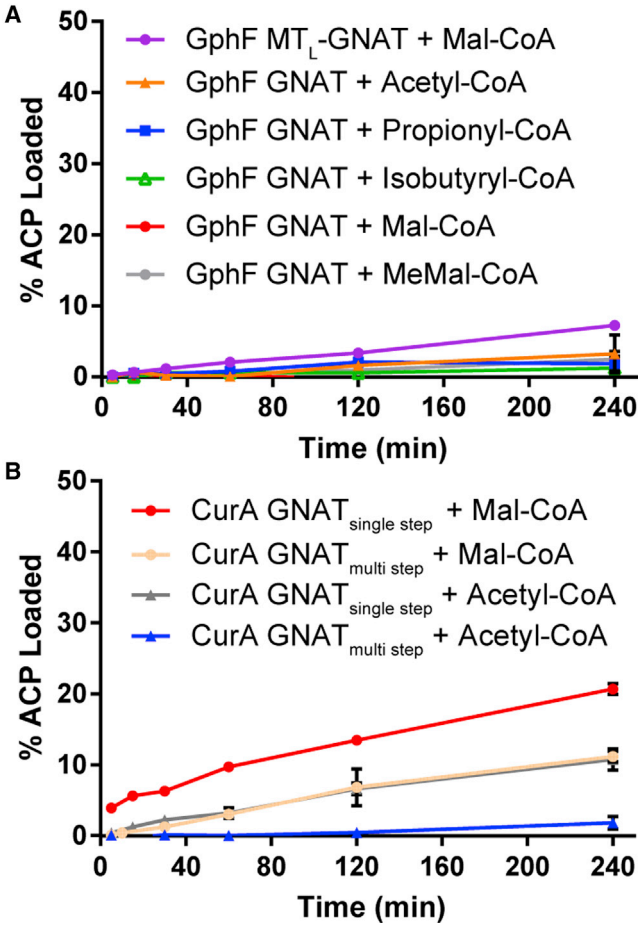


Figure 7. GphF and CurA GNAT Acyl Transfer Assays

(A) Time course (4 h) of 10 μM GphF GNAT and 10 μM GphF MT$_L$-GNAT reactions with acyl-CoAs and AprA holo-ACP. Virtually no acyl group transfer to the ACP was detected with GphF GNAT. GphF MT$_L$-GNAT supported low levels of malonyl transfer.

(B) Time course (4 h) of 10 μM CurA GNAT reactions with Mal-CoA or acetyl-CoA and CurA holo-ACP. The apparent acyl transfer activity of CurA GNAT preparations diminished with additional purification stops, whereas decarboxylation activity was unaffected (see also Figure S8). Activity was monitored via the Ppant ejection assay (Dorrestein et al., 2006; Meluzzi et al., 2008). Error bars represent triplicate experiments and, in some cases, are too small to be observed (see also Figures S6 and S7).

PKS GNAT-like Decarboxylases within the GNAT Superfamily

The PKS GNAT-like decarboxylases, as well as many aminoglycoside N-acetyltransferases and the histone acetyltransferases (HATs), belong to a large protein family (Pfam PF00583, "Acetyltransf_1") (El-Gebali et al., 2019), within the even larger "N-acetyltransferase-like" superfamily (Pfam clan CL0257, Figure S1). GNAT-like domains in polyketide pathways are not the only decarboxylases in the GNAT superfamily. Although studied for decades (Hayaishi, 1955), the catalytic domain of Mal-CoA decarboxylase (MCD) was identified as a member of the GNAT superfamily only when a crystal structure revealed strong similarity to the CurA GNAT-like domain (Figure 1) (Aparicio et al., 2013; Froese et al., 2013). In mammals, MCD modulates levels of Mal-CoA and in some cases can promote the use of MeMal-CoA in fatty acid biosynthesis, leading to tissue-specific production of multimethyl-branched fatty acids (Kim and Kolattukudy, 1978). Disruption of the gene encoding MCD causes the genetic disorder malonic aciduria (Sacksteder et al., 1999). Bacterial genes for MCD are linked to those for a Mal-CoA synthetase and malonate transporter (An and Kim, 1998). Located on a remote island (PF05292, "MCD") within the N-acetyltransferase-like superfamily clan (Figure S1), MCD catalyzes decarboxylation, lacks acyl transfer activity, has very low sequence identity (~10%) to N-acyltransferase members of the GNAT superfamily, and is conserved from bacteria to mammals (An and Kim, 1998; Kim and Kolattukudy, 1978). Despite this distant evolutionary relationship, the GNAT-like domains of PKS initiation modules and the MCD decarboxylases (An and Kim, 1998; Aparicio et al., 2013; Froese et al., 2013) are mechanistically alike. The essential His and Ser/Thr residues are in topologically identical positions on the identical folds of the PKS GNAT-like decarboxylases and MCD catalytic domains (Aparicio et al., 2013; Froese et al., 2013). By reference to the CurA GNAT, analogous His and Ser residues were implicated in MCD decarboxylation (Froese et al., 2013).

Another biosynthetic enzyme for polyketide initiation resides on a distant branch (PF13523, "Acetyltransf_8") of the N-acetyltransferase-like superfamily tree (clan CL0257) (Figure S1). The malonyl- and methylmalonyl-CoA decarboxylase EryM from *Saccharopolyspora erythraea* (also known as SACE_1304 or Mcd [Lazos et al., 2010; Oliynyk et al., 2007; Robbel et al., 2011]) is essential for the production of both the macrolide antibiotic erythromycin (Hsieh and Kolattukudy, 1994) and the siderophore erythrochelin (Lazos et al., 2010). EryM is encoded in an otherwise defunct biosynthetic gene cluster (*nrps1*) that is remote from both the erythromycin and erythrochelin clusters. The multifunctional EryM decarboxylates methylmalonyl-CoA to generate a propionyl-CoA starter unit for erythromycin biosynthesis (Hsieh and Kolattukudy, 1994). Remarkably, in erythrochelin biosynthesis, EryM both decarboxylates malonyl-CoA and transfers acetyl to δ-N-hydroxy-L-ornithine (Lazos et al., 2010; Robbel et al., 2011). The C-terminal domain of EryM was recognized as a member of the GNAT superfamily based on sequence similarity to other GNATs (Lazos et al., 2010; Neuwald and Landsman, 1997; Robbel et al., 2011). The sequence identity to PKS GNAT-like domains (8%–15%) is too low for reliable homology modeling, but it is highly similar (36% identity) to the structurally characterized N-acyltransferase MbtK (Rv1347c) for biosynthesis of the mycobactin siderophore in *Mycobacterium tuberculosis* (Card et al., 2005; Krithika et al., 2006). MbtK and the GNAT-like domain of EryM do not contain His and Ser/Thr residues in analogous locations to the amino acids essential for catalysis by the PKS GNAT-like decarboxylases or MCD. Instead MbtK, EryM, and a number of unannotated EryM homologs contain conserved His and Glu(Asp) amino acids (Figure S2). In MbtK, these amino acids are critical for long-chain fatty acyl transfer from ACP to lysine N^6 (Frankel and Blanchard, 2008; Krithika et al., 2006; McMahon et al., 2012). The His and Glu(Asp) map to the V-shaped Ppant binding pocket, but are on the opposite side of the central β sheet relative to the catalytic resides of

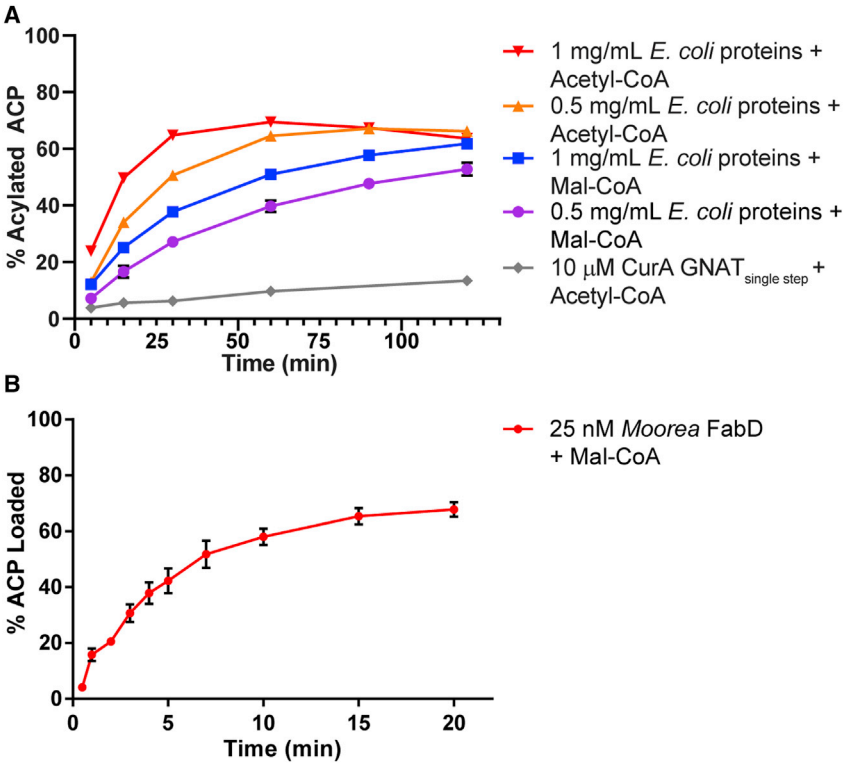


Figure 8. Time Courses for acyltransfer by the *E. coli* Protein Mixture in Reactions with CurA Holo-ACP, and acetyl-CoA or Mal-CoA and *M. bouillonii* FabD Reactions with CurA holo-ACP and Mal-CoA

(A) Acyltransfer by the *E. coli* mixture was concentration dependent. Acetyl transfer from single-step purified CurA GNAT (Figure 7B) is shown for comparison (see also Figure S8).

(B) 25 nM *M. bouillonii* FabD rapidly transferred malonyl from CoA to CurA ACP (20 min time course) at a 1:4,000 molar ratio of FabD to ACP (see also Figure S7). Error bars represent triplicate experiments and, in some cases, are too small to be visible.

the PKS GNAT-like decarboxylases and MCD. Therefore, EryM relies on alternative decarboxylation mechanism. This is a remarkable example of independent evolution of catalytic activities on a fold.

The enormous N-acetyltransferase-like superfamily (>280,000 members), including the HATs of epigenetic regulation and the aminoglycoside N-acetyltransferases, is classically recognized for the GNAT fold and for acetylation of a diverse set of substrates from an acetyl-CoA donor (Dyda et al., 2000). Our work on GNAT-like enzymes in polyketide biosynthesis illuminates additional chemical reactivity within the superfamily. In polyketide initiation modules, the GNAT fold has been co-opted for selective decarboxylation of Mal-ACP, MeMal-ACP, or Me_2Mal-ACP with a strict substrate selectivity that ensures synthesis of the correct natural product. The identical catalytic machinery is employed by these PKS GNAT-like decarboxylases and by MCD, although their sequences are highly diverged. Interestingly, another branch of the GNAT superfamily includes the bifunctional decarboxylase/acyltransferase EryM of erythromycin and erythrochelin biosynthesis (Hsieh and Kolattukudy, 1994) and the fatty acyltransferase of mycobactin biosynthesis. EryM lacks the residues required for decarboxylation by MCD and the PKS GNAT-like decarboxylases, indicating that the superfamily harbors members with an alternative decarboxylation mechanism. Additional uncharacterized decarboxylase members of the GNAT superfamily may currently be misannotated as acyltransferases. Overall, the common feature of many GNAT acyltransferases and decarboxylases is the ability to bind the pantetheine moiety of CoA or ACP at the characteristic V-shaped Ppant binding pocket in the central β sheet and to stabilize a thioester enolate common to the acyltransfer and decarboxylation reactions on CoA- or ACP-linked substrates. The broad substrate scope and varied catalytic functions of the GNAT fold provide a fascinating example of divergent enzymatic evolution and provoke interest into the evolutionary origin of the superfamily.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Construct Design
  - Protein Expression and Purification
  - Production of Acyl-ACPs
  - Enrichment of Non-specific Ni-NTA Binders
  - Decarboxylation Enzyme Assays
  - Acyl Transfer Enzyme Assays
  - LC-MS Analysis
  - Protein Crystallization and Structure Determination
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILIBILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.str.2019.11.004.

ACKNOWLEDGMENTS

We thank R. Taylor (University of Notre Dame) for the *gphF* DNA and R. Sturgis for assistance with cloning. This work was supported by NIH grants DK042303 to J.L.S. and CA108874 to D.H.S., W.H.G., and J.L.S. M.A.S. was supported by the NIH Cellular Biotechnology Training Program (GM008353) and a predoctoral fellowship from the University of Michigan Rackham Graduate School. A.P.S. was supported by a predoctoral fellowship from an NIH Molecular Biophysics Training Program (GM008270). GM/CA@APS is supported by the NIH National Institute of General Medical Sciences (AGM-12006) and National Cancer Institute (ACB-12002).

AUTHOR CONTRIBUTIONS

M.A.S., A.P.S., and J.L.S. designed the research. M.A.S., C.L.T., Q.D., A.P.S., and Z.K. performed the research. M.A.S., C.L.T., A.P.S., W.P.G., D.H.S., and J.L.S. analyzed the data. M.A.S. and J.L.S. wrote the paper with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 21, 2019
Revised: October 6, 2019
Accepted: November 8, 2019
Published: November 27, 2019

REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

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.

An, J.H., and Kim, Y.S. (1998). A gene cluster encoding malonyl-CoA decarboxylase (MatA), malonyl-CoA synthetase (MatB) and a putative dicarboxylate carrier protein (MatC) in *Rhizobium trifolii*—cloning, sequencing, and expression of the enzymes in *Escherichia coli*. Eur. J. Biochem. *257*, 395–402.

Aparicio, D., Perez-Luque, R., Carpena, X., Diaz, M., Ferrer, J.C., Loewen, P.C., and Fita, I. (2013). Structural asymmetry and disulfide bridges among subunits modulate the activity of human malonyl-CoA decarboxylase. J. Biol. Chem. *288*, 11907–11919.

Card, G.L., Peterson, N.A., Smith, C.A., Rupp, B., Schick, B.M., and Baker, E.N. (2005). The crystal structure of Rv1347c, a putative antibiotic resistance protein from *Mycobacterium tuberculosis*, reveals a GCN5-related fold and suggests an alternative function in siderophore biosynthesis. J. Biol. Chem. *280*, 13978–13986.

Chang, Z., Sitachitta, N., Rossi, J.V., Roberts, M.A., Flatt, P.M., Jia, J., Sherman, D.H., and Gerwick, W.H. (2004). Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. J. Nat. Prod. *67*, 1356–1367.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

Chun, S.W., Hinze, M.E., Skiba, M.A., and Narayan, A.R.H. (2018). Chemistry of a unique polyketide-like synthase. J. Am. Chem. Soc. *140*, 2430–2433.

Couture, J.F., and Trievel, R.C. (2006). Histone-modifying enzymes: encrypting an enigmatic epigenetic code. Curr. Opin. Struct. Biol. *16*, 753–760.

Dorrestein, P.C., Bumpus, S.B., Calderone, C.T., Garneau-Tsodikova, S., Aron, Z.D., Straight, P.D., Kolter, R., Walsh, C.T., and Kelleher, N.L. (2006). Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry. Biochemistry *45*, 12756–12766.

Dutnall, R.N., Tafrov, S.T., Sternglanz, R., and Ramakrishnan, V. (1998). Structure of the histone acetyltransferase Hat1: a paradigm for the GCN5-related N-acetyltransferase superfamily. Cell *94*, 427–438.

Dyda, F., Klein, D.C., and Hickman, A.B. (2000). GCN5-related N-acetyltransferases: a structural overview. Annu. Rev. Biophys. Biomol. Struct. *29*, 81–103.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019). The Pfam protein families database in 2019. Nucleic Acids Res. *47*, D427–D432.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Farazi, T.A., Waksman, G., and Gordon, J.I. (2001). Structures of *Saccharomyces cerevisiae* N-myristoyltransferase with bound myristoylCoA and peptide provide insights about substrate recognition and catalysis. Biochemistry *40*, 6335–6343.

Favrot, L., Blanchard, J.S., and Vergnolle, O. (2016). Bacterial GCN5-related N-acetyltransferases: from resistance to regulation. Biochemistry *55*, 989–1002.

Florova, G., Kazanina, G., and Reynolds, K.A. (2002). Enzymes involved in fatty acid and polyketide biosynthesis in *Streptomyces glaucescens*: role of FabH and FabD and their acyl carrier protein specificity. Biochemistry *41*, 10462–10471.

Frankel, B.A., and Blanchard, J.S. (2008). Mechanistic analysis of *Mycobacterium tuberculosis* Rv1347c, a lysine N^{ϵ}-acyltransferase involved in mycobactin biosynthesis. Arch. Biochem. Biophys. *477*, 259–266.

Froese, D.S., Forouhar, F., Tran, T.H., Vollmar, M., Kim, Y.S., Lew, S., Neely, H., Seetharaman, J., Shen, Y., Xiao, R., et al. (2013). Crystal structures of malonyl-coenzyme A decarboxylase provide insights into its catalytic mechanism and disease-causing mutations. Structure *21*, 1182–1192.

Gould, T.A., Schweizer, H.P., and Churchill, M.E. (2004). Structure of the *Pseudomonas aeruginosa* acyl-homoserinelactone synthase LasI. Mol. Microbiol. *53*, 1135–1146.

Grindberg, R.V., Ishoey, T., Brinza, D., Esquenazi, E., Coates, R.C., Liu, W.T., Gerwick, L., Dorrestein, P.C., Pevzner, P., Lasken, R., et al. (2011). Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. PLoS One *6*, e18565.

Gu, L., Geders, T.W., Wang, B., Gerwick, W.H., Hakansson, K., Smith, J.L., and Sherman, D.H. (2007). GNAT-like strategy for polyketide chain initiation. Science *318*, 970–974.

Hayaishi, O. (1955). Enzymatic decarboxylation of malonic acid. J. Biol. Chem. *215*, 125–136.

Hickman, A.B., Namboodiri, M.A., Klein, D.C., and Dyda, F. (1999). The structural basis of ordered substrate binding by serotonin N-acetyltransferase: enzyme complex at 1.8 Å resolution with a bisubstrate analog. Cell *97*, 361–369.

Hsieh, Y.J., and Kolattukudy, P.E. (1994). Inhibition of erythromycin synthesis by disruption of malonyl-coenzyme A decarboxylase gene *eryM* in *Saccharopolyspora erythraea*. J. Bacteriol. *176*, 714–724.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Kampa, A., Gagunashvili, A.N., Gulder, T.A., Morinaka, B.I., Daolio, C., Godejohann, M., Miao, V.P., Piel, J., and Andresson, O. (2013). Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses. Proc. Natl. Acad. Sci. U S A *110*, E3129–E3137.

Kellmann, R., Mihali, T.K., Jeon, Y.J., Pickford, R., Pomati, F., and Neilan, B.A. (2008). Biosynthetic intermediate analysis and functional homology reveal a saxitoxin gene cluster in cyanobacteria. Appl. Environ. Microbiol. *74*, 4044–4053.

Kim, Y.S., and Kolattukudy, P.E. (1978). Malonyl-CoA decarboxylase from the uropygial gland of waterfowl: purification, properties, immunological

comparison, and role in regulating the synthesis of multimethyl-branched fatty acids. Arch. Biochem. Biophys. *190*, 585–597.

Krithika, R., Marathe, U., Saxena, P., Ansari, M.Z., Mohanty, D., and Gokhale, R.S. (2006). A genetic locus required for iron acquisition in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U S A *103*, 2069–2074.

Kumar, P., Koppisch, A.T., Cane, D.E., and Khosla, C. (2003). Enhancing the modularity of the modular polyketide synthases: transacylation in modular polyketide synthases catalyzed by malonyl-CoA: ACP transacylase. J. Am. Chem. Soc. *125*, 14307–14312.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and clustal X version 2.0. Bioinformatics *23*, 2947–2948.

Lazos, O., Tosin, M., Slusarczyk, A.L., Boakes, S., Cortes, J., Sidebottom, P.J., and Leadlay, P.F. (2010). Biosynthesis of the putative siderophore erythrochelin requires unprecedented crosstalk between separate nonribosomal peptide gene clusters. Chem. Biol. *17*, 160–173.

Lombard, J., and Moreira, D. (2011). Early evolution of the biotin-dependent carboxylase family. BMC Evol. Biol. *11*, 232.

Mattheus, W., Gao, L.J., Herdewijn, P., Landuyt, B., Verhaegen, J., Masschelein, J., Volckaert, G., and Lavigne, R. (2010). Isolation and purification of a new kalimantacin/batumin-related polyketide antibiotic and elucidation of its biosynthesis gene cluster. Chem. Biol. *17*, 149–159.

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.

McMahon, M.D., Rush, J.S., and Thomas, M.G. (2012). Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. J. Bacteriol. *194*, 2809–2818.

Meluzzi, D., Zheng, W.H., Hensler, M., Nizet, V., and Dorrestein, P.C. (2008). Top-down mass spectrometry on low-resolution instruments: characterization of phosphopantetheinylated carrier domains in polyketide and non-ribosomal biosynthetic pathways. Bioorg. Med. Chem. Lett. *18*, 3107–3111.

Moebius, N., Ross, C., Scherlach, K., Rohm, B., Roth, M., and Hertweck, C. (2012). Biosynthesis of the respiratory toxin bongkrekic acid in the pathogenic bacterium *Burkholderia gladioli*. Chem. Biol. *19*, 1164–1174.

Neuwald, A.F., and Landsman, D. (1997). GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. Trends Biochem. Sci. *22*, 154–155.

Oliynyk, M., Samborskyy, M., Lester, J.B., Mironenko, T., Scott, N., Dickens, S., Haydock, S.F., and Leadlay, P.F. (2007). Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. Nat. Biotechnol. *25*, 447–453.

Partida-Martinez, L.P., de Looss, C.F., Ishida, K., Ishida, M., Roth, M., Buder, K., and Hertweck, C. (2007). Rhizoxin, the first mycotoxin isolated from the zygomycota, is not a fungal metabolite but is produced by bacterial endosymbionts. Appl. Environ. Microbiol. *73*, 793–797.

Piel, J. (2002). A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. Proc. Natl. Acad. Sci. U S A *99*, 14002–14007.

Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., and Matsunaga, S. (2004a). Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. Proc. Natl. Acad. Sci. U S A *101*, 16222–16227.

Piel, J., Wen, G., Platzer, M., and Hui, D. (2004b). Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase. Chembiochem *5*, 93–98.

Robbel, L., Helmetag, V., Knappe, T.A., and Marahiel, M.A. (2011). Consecutive enzymatic modification of ornithine generates the hydroxamate moieties of the siderophore erythrochelin. Biochemistry *50*, 6073–6080.

Sacksteder, K.A., Morrell, J.C., Wanders, R.J., Matalon, R., and Gould, S.J. (1999). MCD encodes peroxisomal and cytoplasmic forms of malonyl-CoA decarboxylase and is mutated in malonyl-CoA decarboxylase deficiency. J. Biol. Chem. *274*, 24461–24468.

Salah Ud-Din, A.I., Tikhomirova, A., and Roujeinikova, A. (2016). Structure and functional diversity of GCN5-related N-acetyltransferases (GNAT). Int. J. Mol. Sci. *17*, 1018.

Sánchez, C., Du, L., Edwards, D.J., Toney, M.D., and Shen, B. (2001). Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin. Chem. Biol. *8*, 725–738.

Schrodinger. (2015). The PyMOL Molecular Graphics System, Version 1.8 (Schrodinger LLC).

Simunovic, V., Zapp, J., Rachid, S., Krug, D., Meiser, P., and Muller, R. (2006). Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases. Chembiochem *7*, 1206–1220.

Skiba, M.A., Maloney, F.P., Dan, Q., Fraley, A.E., Aldrich, C.C., Smith, J.L., and Brown, W.C. (2018a). PKS-NRPS enzymology and structural biology: considerations in protein production. Methods Enzymol. *604*, 45–88.

Skiba, M.A., Sikkema, A.P., Moss, N.A., Lowell, A.N., Su, M., Sturgis, R.M., Gerwick, L., Gerwick, W.H., Sherman, D.H., and Smith, J.L. (2018b). Biosynthesis of *t*-butyl in apratoxin A: functional analysis and architecture of a PKS loading module. ACS Chem. Biol. *13*, 1640–1650.

Skiba, M.A., Sikkema, A.P., Moss, N.A., Tran, C.L., Sturgis, R.M., Gerwick, L., Gerwick, W.H., Sherman, D.H., and Smith, J.L. (2017). A mononuclear iron-dependent methyltransferase catalyzes initial steps in assembly of the apratoxin A polyketide starter unit. ACS Chem. Biol. *12*, 3039–3048.

Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F.R., and Donnelly, M.I. (2002). A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. Protein Expr. Purif. *25*, 8–15.

Sudek, S., Lopanik, N.B., Waggoner, L.E., Hildebrand, M., Anderson, C., Liu, H., Patel, A., Sherman, D.H., and Haygood, M.G. (2007). Identification of the putative bryostatin polyketide synthase gene cluster from "*Candidatus endobugula sertula*", the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. J. Nat. Prod. *70*, 67–74.

Tsay, J.T., Oh, W., Larson, T.J., Jackowski, S., and Rock, C.O. (1992). Isolation and characterization of the beta-ketoacyl-acyl carrier protein synthase III gene (fabH) from *Escherichia coli* K-12. J. Biol. Chem. *267*, 6807–6814.

Van Wagoner, R.M., and Clardy, J. (2006). FeeM, an N-acyl amino acid synthase from an uncultured soil microbe: structure, mechanism, and acyl carrier protein binding. Structure *14*, 1425–1435.

Vetting, M.W., Errey, J.C., and Blanchard, J.S. (2008). Rv0802c from *Mycobacterium tuberculosis*: the first structure of a succinyltransferase with the GNAT fold. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. *64*, 978–985.

Vetting, M.W., S de Carvalho, L.P., Yu, M., Hegde, S.S., Magnet, S., Roderick, S.L., and Blanchard, J.S. (2005). Structure and functions of the GNAT superfamily of acetyltransferases. Arch. Biochem. Biophys. *433*, 212–226.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Watson, W.T., Minogue, T.D., Val, D.L., von Bodman, S.B., and Churchill, M.E. (2002). Structural basis and specificity of acyl-homoserine lactone signal production in bacterial quorum sensing. Mol. Cell *9*, 685–694.

Wolf, E., Vassilev, A., Makino, Y., Sali, A., Nakatani, Y., and Burley, S.K. (1998). Crystal structure of a GCN5-related N-acetyltransferase: *Serratia marcescens* aminoglycoside 3-N-acetyltransferase. Cell *94*, 439–449.

Young, J., Stevens, D.C., Carmichael, R., Tan, J., Rachid, S., Boddy, C.N., Muller, R., and Taylor, R.E. (2013). Elucidation of gephyronic acid biosynthetic pathway revealed unexpected SAM-dependent methylations. J. Nat. Prod. *76*, 2269–2276.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Escherichia coli BL21(DE3)	Novagen	Cat# 70235
Chemicals, Peptides, and Recombinant Proteins		
GphF GNAT	This paper	residues 498-705, GenBank: KF479198.1
GphF MT1-GNAT	Skiba et al., 2017	residues 2-696, GenBank: KF479198.1
AprA ACP	Skiba et al., 2017	residues 1058-1138, GenBank: WP_075900460
CurA GNAT	Gu et al., 2007	residues 219-439, GenBank: AEE88289.1
CurA ACP	This paper	residues 444-521, GenBank: AEE88289.1
AprA MT1-ΨGNAT	Skiba et al., 2017	residues 2-629, GenBank: WP_075900460
S. verticillus phosphopantetheinyl transferse	Sánchez et al., 2001	GenBank: AF210311
S-adenosymethionine	Carbosynth	Cat# NA04017
$Fe(NH_4)_2(SO_4)_2$	Fisher Scientific	Cat# I77-500
Acetyl-CoA	CoALA Biosciences	Cat# SKU AC01
Malonyl-CoA	CoALA Biosciences	Cat# SKU MC01
Methylmalonyl-CoA	CoALA Biosciences	Cat# SKU MM01
Propionyl-CoA	CoALA Biosciences	Cat# SKU PC01
Isobutyryl-CoA	CoALA Biosciences	Cat# SKU IC01
PEG 3350	Hampton Research	Cat# HR2-591
Ammonium Acetate	Mallinckrodt Chemicals	Cat# 3272-04
Bis-Tris	Fisher Scientific	Cat# BP301-100
HisTrap column	GE Healthcare	Cat# 17524801
Ni-NTA agarose	Qiagen	Cat# 30210
HiLoad 16/60 Superdex S75	GE Healthcare	Cat# GE28-9893-33
HiTrap Q anion exchange column	GE Healthcare	Cat# 17115401
HiLoad 16/60 Superdex S200	GE Healthcare	Cat# GE28-9893-35
Deposited Data		
GphF GNAT	This Paper	PDB: 6MFC
GphF GNAT in complex with isobutyryl-CoA	This Paper	PDB: 6MFD
CurA GNAT	Gu et al., 2007	PDB: 2REE
CurA GNAT in complex with malonyl-CoA	Gu et al., 2007	PDB: 2REF
Recombinant DNA		
Plasmid:pMCSG7	Stols et al., 2002	N/A
Software and Algorithms		
MassHunter Qualitative Analysis	Agilent	https://www.agilent.com/en/products/software-informatics/masshunter-suite/masshunter/masshunter-software
XDS	Kabsch, 2010	http://xds.mpimf-heidelberg.mpg.de/html_doc/downloading.html
Phenix Software Suite	Adams et al., 2010	https://www.phenix-online.org/
Coot	Emsley and Cowtan, 2004	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
PyMol	Schrodinger, 2015	https://pymol.org/2/
Jalview	Waterhouse et al., 2009	https://www.jalview.org/
GraphPad Prism	GraphPad Software, La Jolla California USA	https://www.graphpad.com

LEAD CONTACT AND MATERIALS AVAILABILITY

Requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Janet Smith (JanetSmith@umich.edu). All plasmids generated in this study are available from the Lead Contact with a Materials Transfer Agreement.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

E. coli BL21(DE3) was used for heterologous expression of all proteins in this study.

METHOD DETAILS

Construct Design

DNA encoding the GphF GNAT was amplified from a partial *gphF* clone provided by Richard Taylor (Notre Dame University) and inserted into pMCSG7 (Stols et al., 2002) by ligation independent cloning (LIC) to create the expression plasmid for GphF GNAT (residues 498-705, GenBank: KF479198.1; pMAS/APS340). A cDNA encoding the CurA ACP (residues 444-521, GenBank: AEE88289.1) was amplified from a cosmid library (Chang et al., 2004) and inserted into pMCSG7 via LIC to create pMAS/APS442. All *gphF* site-directed mutants (Table S2) were introduced into pMAS340 using the QuickChange protocol (Stratagene). All primers are listed in Table S1. All constructs and mutations were verified by Sanger sequencing at the University of Michigan DNA Sequencing Core.

Protein Expression and Purification

Escherichia coli strain BL21(DE3) was transformed with plasmids encoding GphF GNAT and CurA GNAT (residues 219-439, GenBank: AEE88289.1 in pMCSG7) (Gu et al., 2007). Transformed cells were grown in 0.5 L of TB media at 37°C supplemented with 100 μg mL^{-1} ampicillin to an OD_{600}=1-2, cooled to 20°C for 1 hr, and induced with 200 μM IPTG for 18 hr. Cell pellets were resuspended in 35 mL Tris buffer A (50 mM Tris pH 7.4, 300 mM NaCl, 10% (v/v) glycerol, 20 mM imidazole), 0.1 mg mL^{-1} lysozyme, 0.05 mg mL^{-1} DNase, and 2 mM $MgCl_2$, incubated on ice for 30 min, lysed by sonication, and cleared by centrifugation (38,760 x g, 15-30 min, 4°C). The supernatant was filtered and loaded onto a 5 mL HisTrap column (GE Healthcare) or 5 mL of Ni-NTA resin (Qiagen) equilibrated with Tris buffer A. After washing with 10 column volumes of Tris buffer A, proteins were eluted with a 5-100% gradient of Tris buffer B (50 mM Tris 7.4, 300 mM NaCl, 10% glycerol, 400 mM imidazole) over 10 column volumes. Single-step purified CurA GNAT was dialyzed into Tris Buffer A for acyl transfer assays. Proteins used for decarboxylation assays and crystallization were further purified by size exclusion chromatography (HiLoad 16/60 Superdex S75) in Tris buffer C (50 mM Tris pH 7.4, 150 mM NaCl, 10% glycerol).

GphF GNAT used for acyl transfer assays was dialyzed into 50 mM BisTris pH 6.5, 20 mM NaCl, 10% glycerol (BisTris buffer A) and loaded on a HiTrap Q anion exchange column (GE Healthcare) equilibrated with BisTris buffer A. GphF GNAT was collected from the Q column flow through and buffer exchanged into Tris buffer C via size exclusion chromatography (HiLoad 16/60 Superdex S75).

A multistep purification was used for CurA GNAT intended for acyl transfer assays. Cell pellets were resuspended in 5 mL Tris buffer D (100 mM Tris pH 7.9, 300 mM NaCl, 10% glycerol, 15 mM imidazole) per 1 g cell pellet with 4 mg DNase, 10 mg lysozyme, and 4 mM $MgCl_2$ and incubated on ice for 30 min. Cells were lysed by three passes through an Avestin EmulsiFlex-C3 homogenizer and cleared by centrifugation (30,000 x g, 30 min, 4°C). The supernatant was filtered and loaded onto a 5 mL HisTrap column equilibrated with Tris buffer D. Protein was eluted with a gradient of 30-300 mM imidazole. The His-tag was cleaved by overnight incubation with tobacco etch virus (TEV) protease (1:30 ratio of TEV protease: CurA GNAT with 2 mM DTT) and dialyzed into Tris buffer D. The digested protein was passed over a second 5 mL HisTrap column. Tag-free CurA GNAT was collected from the flow through and further purified by size exclusion chromatography (HiLoad 16/60 Superdex S200) in Tris buffer D.

GphF MT1-GNAT (residues 2-696, GenBank: KF479198.1) used for acyl transfer assays was purified as described previously (Skiba et al., 2017) and then diluted into Hepes buffer A (50 mM Hepes pH 7.4, 15 mM NaCl, 10% glycerol). GphF MT1-GNAT was then loaded onto a HiTrap Q Column equilibrated with Hepes buffer A, washed with 10 column volumes of Hepes buffer A and eluted with a 0.15-1 M linear gradient of NaCl over 20 column volumes. GphF MT-GNAT was buffer exchanged into Hepes buffer B (50 mM Hepes pH 7.4, 150 mM NaCl, 10% glycerol) via size exclusion chromatography (HiLoad 16/60 Superdex S200).

AprA apo-ACP (residues 1058-1138, GenBank: WP_075900460) was produced and purified as described previously (Skiba et al., 2017). A trace metals mix was included in the growth medium to inhibit addition of the phosphopantetheinyl post translational modification by endogenous *E. coli* enzymes (Skiba et al., 2018a). CurA ACP (residues 444-519, GenBank: AEE88289.1) was produced and purified identically to AprA ACP. *M. bouillonii* FabD (residues 1-298, GenBank: WP_075905309.1) (Skiba et al., 2018b) and AprA MT1-ΨGNAT (residues 2-629, GenBank: WP_075900460) (Skiba et al., 2017) were produced and purified as described previously.

Production of Acyl-ACPs

AprA and CurA holo-, malonyl- (Mal-), and methylmalonyl- (MeMal-)ACPs were produced by incubating 180 μM apo-ACP with 20 μM *Streptomyces verticillus* phosphopantetheinyl transferase (SVP) (Sánchez et al., 2001), and ~0.65 mM CoA, Mal-CoA, or MeMal-CoA

in 50 mM Tris pH 7.4, 150 mM NaCl, 10% glycerol, 20 mM $MgCl_2$ for 4 hr at 30°C. The ACP was purified from the reaction mixture by size exclusion chromatography (HiLoad 16/60 Superdex S75) equilibrated with Tris buffer E (100 mM Tris pH 7.4, 250 mM NaCl, 5% glycerol, 5 mM tris(2-carboxyethyl)phosphine (TCEP).

As dimethylmalonyl (Me_2Mal)-CoA is not commercially available, AprA MT1 (Skiba et al., 2017) was used to convert the AprA and CurA MeMal-ACPs to the Me_2Mal-ACPs. Reaction mixtures for AprA ACP (~3 mL total in 130 μL aliqots) containing 300 μM AprA MeMal-ACP, 150 μM AprA MT1-ΨGNAT and 6 mM *S*-adenosylmethionine (SAM) in 50 mM Hepes pH 7.4, 150 mM NaCl, 3 mM $Fe(NH_4)_2(SO_4)_2$ were incubated 5.5 hr at 30°C. Reaction mixtures for CurA ACP (~4 mL total in 50 μL aliqots) containing 150 μM CurA MeMal-ACP, 150 μM AprA MT1-ΨGNAT and 6 mM SAM in 50 mM Hepes pH 7.4, 150 mM NaCl, 3 mM $Fe(NH_4)_2(SO_4)_2$ were incubated for 6 hr at 30°C. Me_2Mal-ACPs were purified from the reaction mixtures by size exclusion chromatography (HiLoad 16/60 Superdex S75) equilibrated with Tris buffer E. Intact protein mass spectra coupled with the Ppant ejection assay (Dorrestein et al., 2006; Meluzzi et al., 2008) was used to validate the conversion of MeMal-ACP to Me_2Mal-ACP.

Enrichment of Non-specific Ni-NTA Binders

Empty pMCSG7 expression vector (Stols et al., 2002) was transformed into *E. coli* BL21(DE3), expressed and purified as described for the GphF and CurA GNAT above. His elution fractions that would typically contain the GNAT were collected, concentrated 20 fold in a 3 kDa cutoff concentrator and dialyzed into Tris Buffer A to remove imidazole.

Decarboxylation Enzyme Assays

As GphF ACP_L could not be produced in soluble form, AprA ACP_L from the apratoxin A biosynthetic pathway (Grindberg et al., 2011) was used as a surrogate ACP. Reaction mixtures (80-100 μL) containing 100 μM AprA Mal-, MeMal-, or Me_2Mal-ACP and 10 μM GphF GNAT in 50 mM HEPES pH 7.4, 150 mM NaCl were incubated at 30°C. 10 μL of the reaction mixture was collected at various time points and quenched with 1% (v/v) formic acid. 0.25 μL of reaction mixtures were subjected to LC/MS analysis.

Reaction mixtures (10 μL) containing 100 μM AprA acyl-ACP and 10 μM GphF GNAT variants in 50 mM HEPES pH 7.4, 150 mM NaCl were incubated at 30°C for 5 min (Me_2Mal-ACP), 8 hr (MeMal-ACP), or 24 hr (Mal-ACP). Reactions were quenched with 1% (v/v) formic acid. 0.25 μL of reaction mixtures were subjected to LC/MS analysis.

Reaction mixtures (80-100 μL) containing 100 μM CurA Mal-, MeMal-, or Me_2Mal-ACP and 10 μM CurA GNAT in 50 mM HEPES pH 7.4, 150 mM NaCl were incubated at 30°C. 10 μL of the reaction mixture was collected at various time points and quenched with 1% (v/v) formic acid. 0.1 μL of reaction mixtures were subjected to LC/MS analysis.

Reaction mixtures (10 μL) containing 100 μM CurA acyl-ACP and 10 μM CurA GNAT variants in 50 mM HEPES pH 7.4, 150 mM NaCl were incubated 6 hr (Me_2Mal-ACP), 7.5 min (MeMal-ACP), or 3 min (Mal-ACP). Reactions were quenched with 1% (v/v) formic acid. 0.1 μL of reaction mixtures were subjected to LC/MS analysis.

Acyl Transfer Enzyme Assays

Acyl transfer reaction mixtures (70 μL) containing 100 μM AprA or CurA holo-ACP, 10 μM CurA GNAT or GphF GNAT, 0.85 mM of Mal-, MeMal-, acetyl-, propionyl-, or isobutyryl-CoA in 50 mM Hepes pH 7.4, 150 mM NaCl were incubated at 30°C. At various time points 10 μL aliquots were removed and quenched with 1% (v/v) formic acid. 0.25 μL of GphF GNAT reactions and 0.1 μL of CurA GNAT reactions were used for LC/MS analysis.

GphF MT_L-GNAT acyl transfer reaction mixtures (70 μL) containing 100 μM AprA holo-ACP, 10 μM GphF MT_L-GNAT, 0.85 mM of Mal-CoA, 0.54 mM SAM, 50 mM Hepes pH 7.4, 150 mM NaCl, 0.5 mM $FeH_8N_2O_8S_2$ were incubated at 30°C. At various time points 10 μL aliquots were removed and quenched with 1% (v/v) formic acid. 0.25 μL of the reaction mixtures were subjected to LC/MS analysis.

M. bouillonii FabD reaction mixtures (120 μL) containing 100 μM CurA holo-ACP, 25 nM FabD, and 0.85 mM Mal-CoA in 50 mM Hepes 7.4, 150 mM NaCl were incubated at 30°C. At various time points 10 μL aliquots were removed and quenched with 1% (v/v) formic acid. 0.1 μL of the reaction mixtures were subjected to LC/MS analysis.

Reaction mixtures (120 μL) with *E. coli* proteins from the mock purification contained 0.5 or 1 mg mL^{-1} *E. coli* proteins, 100 μM CurA holo-ACP, and 0.87 mM of Mal-CoA or acetyl-CoA in 50 mM Hepes 7.5, 150 mM NaCl and were incubated at 30°C. 10 μL of the reaction mixture was collected at various time points and quenched with 1% (v/v) formic acid. 0.1 μL of the reaction mixture was subjected to LC/MS analysis.

LC-MS Analysis

An Agilent Q-TOF 6545 was used to analyze reaction mixtures using the phosphopantetheine (Ppant) ejection method (Dorrestein et al., 2006; Meluzzi et al., 2008). Samples underwent reverse phase HPLC (Phenomenex Aeris widepore C4 column 3.6 μM, 50 x 2.10 mm) in H_2O with 0.2% (v/v) formic acid at a flow rate of 0.5 mL min^{-1}. Protein was eluted over a 4 min linear gradient of 5-100% acetonitrile with 0.2% (v/v) formic acid. Samples were subjected to the following conditions: fragmentor voltage, 225 V; skimmer voltage, 25 V; nozzle voltage, 1000 V; sheath gas temperature, 350°C; drying gas temperature, 325°C. Data were processed using MassHunter Qualitative Analysis Software (Agilent). The relative abundances of Ppant ejection fragments for substrates and products were used to calculate the percent of total ACP species. In acyl transfer reactions with active GNAT domains, abundances of intact and decarboxylated acyl-ACP species were combined to calculate percent acyl group transfer.

Protein Crystallization and Structure Determination

GphF GNAT (residues 498-705, with additional N-terminal His-tag) was crystallized at 20°C by sitting drop vapor diffusion from a 2:1 μL mixture of protein stock (20.7 mg mL^{-1} GphF GNAT in Tris buffer C) and reservoir solution (30 - 35% PEG 3350, 0.23 - 0.30 M ammonium acetate, 0.10 M Bis-tris HCl pH 5.5). Crystals of thin-blade morphology grew overnight, often in clusters, and could be divided into apparently single crystals for data collection. Crystals were cryoprotected with well solution supplemented with 15% glycerol and flash cooled in liquid N_2. For the isobutyryl-CoA complex, crystals were soaked with 2.5 mM isobutyryl-CoA for 24 hr prior to harvesting.

Diffraction data were collected at 100 K on GM/CA beamline 23ID-B at the Advanced Photon Source (APS) at Argonne National Laboratory (Argonne, IL). Data were processed using XDS (Kabsch, 2010). The thin-blade morphology of the crystals limited the data quality. Diffraction patterns were streaky and moderately anisotropic (diffraction limit for GphF GNAT: d_{min} = 3.3 Å along *a**, and 2.6 Å along *b** and *c**; isobutyryl-CoA complex: d_{min} = 5.0 Å along *a** and 2.8 Å along *b** and *c**). Probing the crystal with a 10-μm beam in raster cells did not identify local regions with better spot shapes or lower anisotropy. The structure of GphF GNAT was solved by molecular replacement using Phaser (McCoy et al., 2007) through the Phenix Software Suite (Adams et al., 2010) with CurA GNAT (2REE) (Gu et al., 2007) as the starting model. Crystals of the isobutyryl complex were isomorphous with crystals of the free enzyme. Chain A contained clear density for isobutyryl-CoA, whereas density in Chain B indicated multiple positions for the CoA product analog. Iterative rounds of model building and refinement were carried out using Coot (Emsley and Cowtan, 2004) and Phenix.refine (Afonine et al., 2012) with automated translation/liberation/screw group selection. Structures were validated with MolProbity (Chen et al., 2010), and structure figures were prepared with PyMol (Schrodinger, 2015). Sequence alignments were prepared using Clustal (Larkin et al., 2007) through Jalview (Waterhouse et al., 2009). The map of GNAT families was generated in the Pfam server (El-Gebali et al., 2019).

QUANTIFICATION AND STATISTICAL ANALYSIS

All biochemical assays were performed in triplicate. Error bars represent standard deviation of the mean.

DATA AND CODE AVAILIBILITY

Atomic coordinates and structure factors have been deposited in the Protein Data Bank for GphF GNAT-like decarboxylase (PDB 6MFC) and the isobutyryl-CoA complex of GphF GNAT-like decarboxylase (PDB 6MFD).

Structure, Volume *28*

Supplemental Information

Repurposing the GNAT Fold in the Initiation of Polyketide Biosynthesis

Meredith A. Skiba, Collin L. Tran, Qingyun Dan, Andrew P. Sikkema, Zachary Klaver, William H. Gerwick, David H. Sherman, and Janet L. Smith

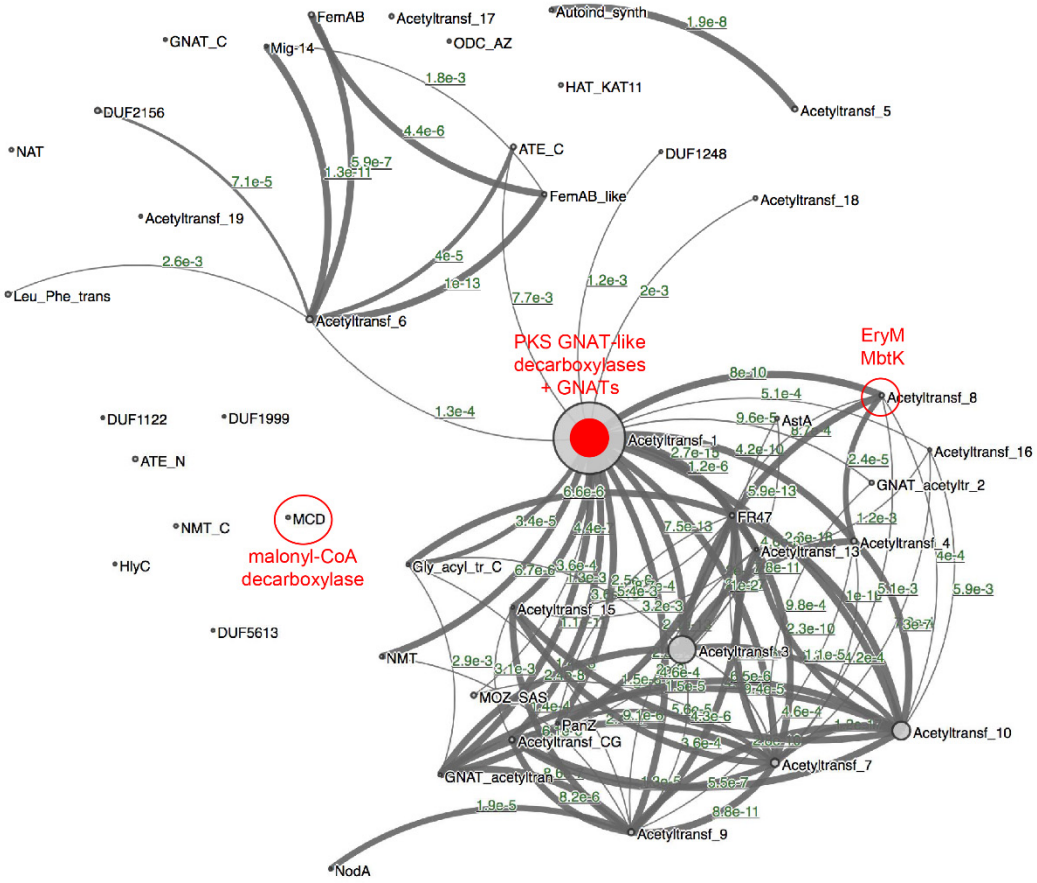


Figure S1. **Family relationships in the *N*-acetyltransferase-like Pfam clan (CL0257, >280,000 sequences).** *Related to Figure 1*. Families discussed in this work are indicated in red. The PKS GNAT-like domains are in the Acetyltransf_1 family (PF00583, >120,000 sequences) in the center of the diagram. The malonyl-CoA decarboxylases are in the isolated MCD family (PF05292, 1442 sequences), and EryM and MbtK belong to the Acetyltransf_8 family (PF13523, 3128 sequences). Image was generated with the Pfam server (https://pfam.xfam.org/).

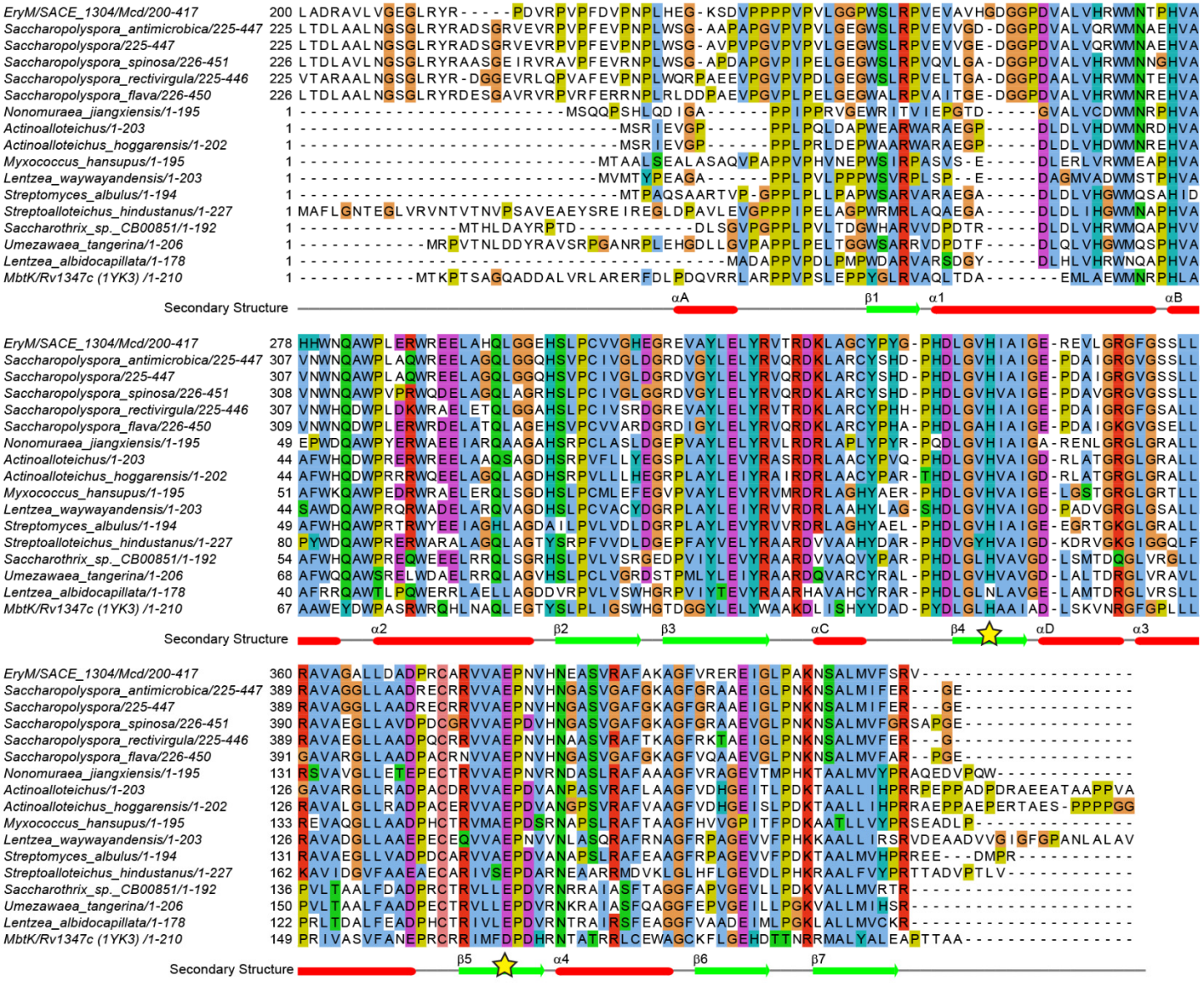


Figure S2. Sequence alignment of the GNAT-like domain of EryM and homologs. *Related to Figure 1*. The secondary structure annotation below the alignment is for the MbtK/Rv1347c crystal structure (PDB 1YK3) (Card et al., 2005). MbtK/Rv1347c catalytic residues, which are conserved in EryM homologs are starred. GenBank accession codes: EryM (*Saccharopolyspora erythrea*), CAM00627.1; *Saccharopolyspora antimicrobica,* WP_093149902.1; *Saccharopolyspora*, WP_093351295.1; *Saccharopolyspora spinosa*, PKW14738.1; *Saccharopolyspora rectivirgula*, WP_051387681.1; *Saccharopolyspora flava*, SFT09105.1; *Nonomuraea jiangxiensis*, WP_090929840.1; *Actinoalloteichus*, WP_075740024.1; *Actinoalloteichus hoggarensis*, WP_093941036.1; *Myxococcus hansupus*, WP_002639252.1; *Lentzea waywayandensis*, WP_093605558.1; *Streptomyces albulus*, WP_064068670.1; *Streptoalloteichus hindustanus*, SHE93970.1; *Saccharothrix sp. CB00851*, WP_083668623.1; *Umezawaea tangerine*, PRY28657.1; *Lentzea albidocapillata*, WP_030481780.1; MtbK/Rv1347c (*Mycobacterium tuberculosis*), WP_003406956.1.

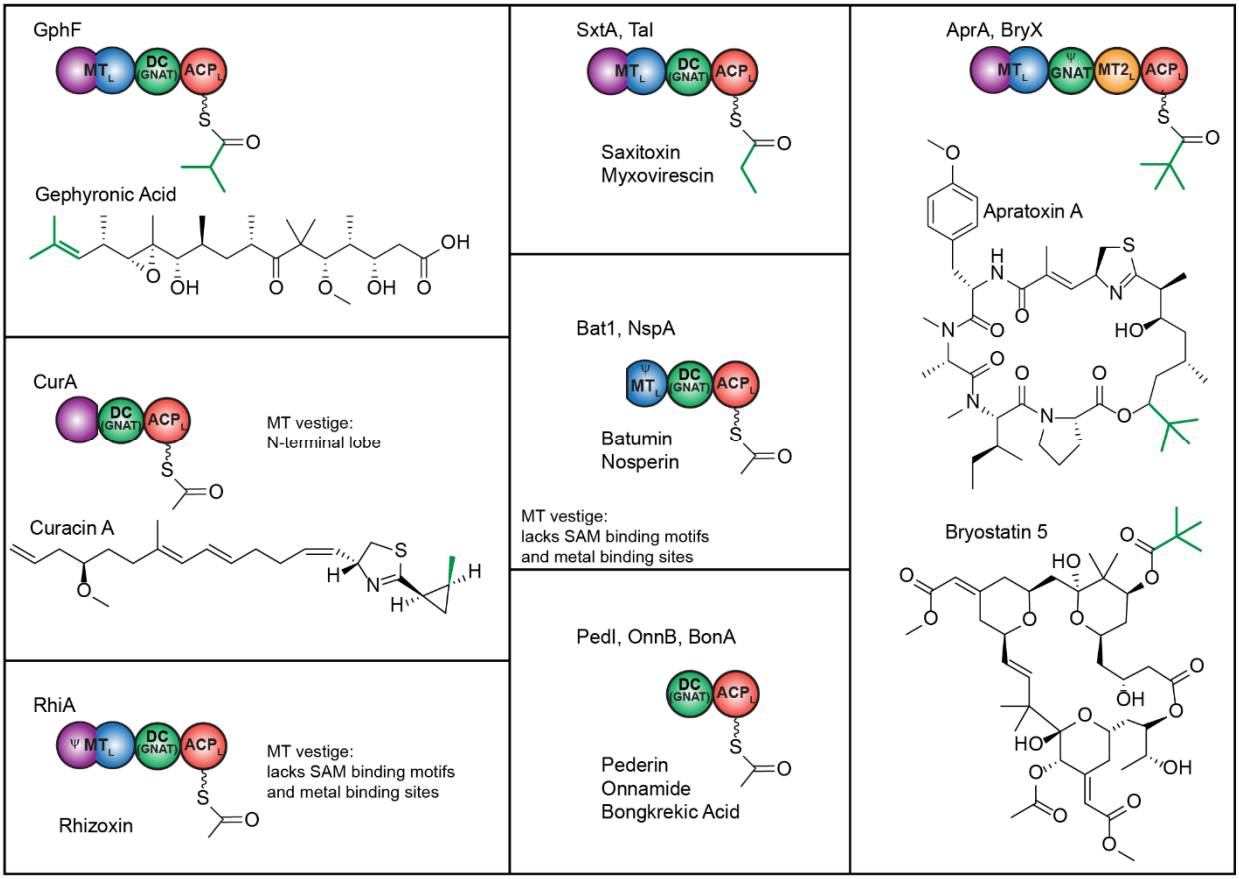


Figure S3. **PKS loading modules with GNAT-like decarboxylase (DC) domains.** *Related to Figure 2*. Domains are colored according to function and homology, and Ψ indicates inactive domains. The metal-dependent MT domain at the N-terminus of most of these loading modules is composed of two sub-domains: an N-terminal lid and a C-terminal MT catalytic core. Some homologs of the GphF MT are inactive: CurA MT lacks a core sub-domain, RhiA MT lacks signature methyltransferase sequence motifs and metal-coordinating residues, and Bat1 and NspA lack a lid sub-domain. As a result, they are expected to load malonyl with decarboxylation by the module's GNAT-like decarboxylase, giving an acetyl starter unit. The MTs of GphF MT, AprA MT1 and BryX MT1 catalyze two methyl transfers to Mal-ACP, whereas the SxtA and TaI MTs catalyze a single methyl transfer. GphF, SxtA and TaI GNAT-like DC domains decarboxylate to produce the isobutyryl or propionyl starter unit. In contrast, the truncated GNAT of AprA and BryX is inactive and the MT2 domain catalyzes a coupled methyl transfer and decarboxylation to produce a pivaloyl starter unit.

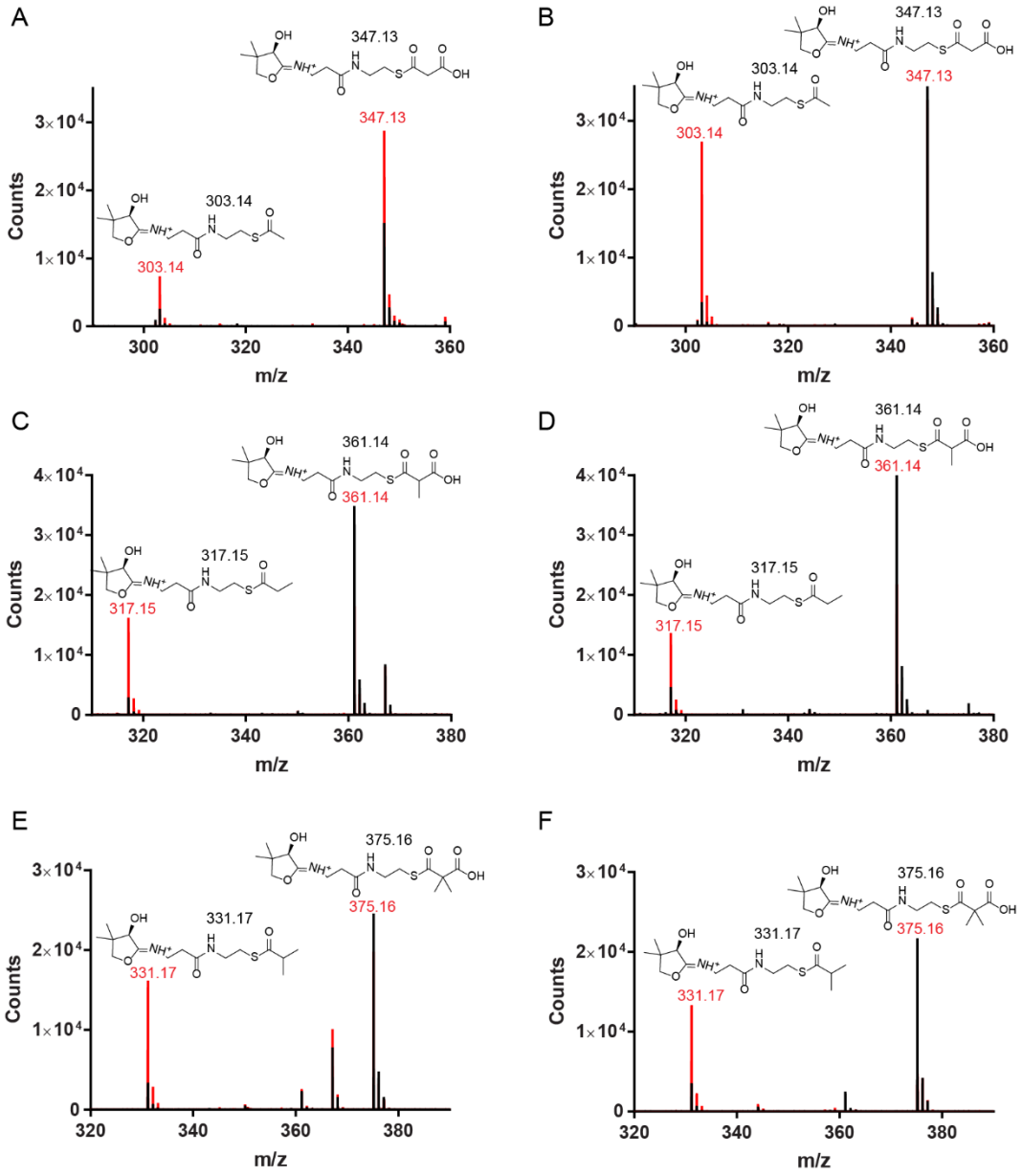


Figure S4. Representative mass spectra of decarboxylation reactions by GNAT-like domains. *Related to Figure 3.*
Decarboxylation of Mal-ACP: **(A)** GphF, 48 hr reaction time; **(B)** CurA, 3 min.
Decarboxylation of MeMal-ACP: **(C)** GphF, 8 hr reaction time; **(D)** CurA, 7.5 min
Decarboxylation of Me_2Mal-ACP: **(E)** GphF, 5 min reaction time; **(F)** CurA, 8 hr.
Spectra of no-enzyme controls are shown in black; reactions are shown in red. Calculated m/z values are in black above structures of the corresponding Ppant ejection fragments, with observed m/z values in red. AprA ACP, used as a surrogate for GphF ACP, had a contaminating species at 367 Da. The no-enzyme control and reaction in **A** had identical ratios of malonyl- and acetyl-Ppant ejection fragments to holo-Ppant ejection fragments.

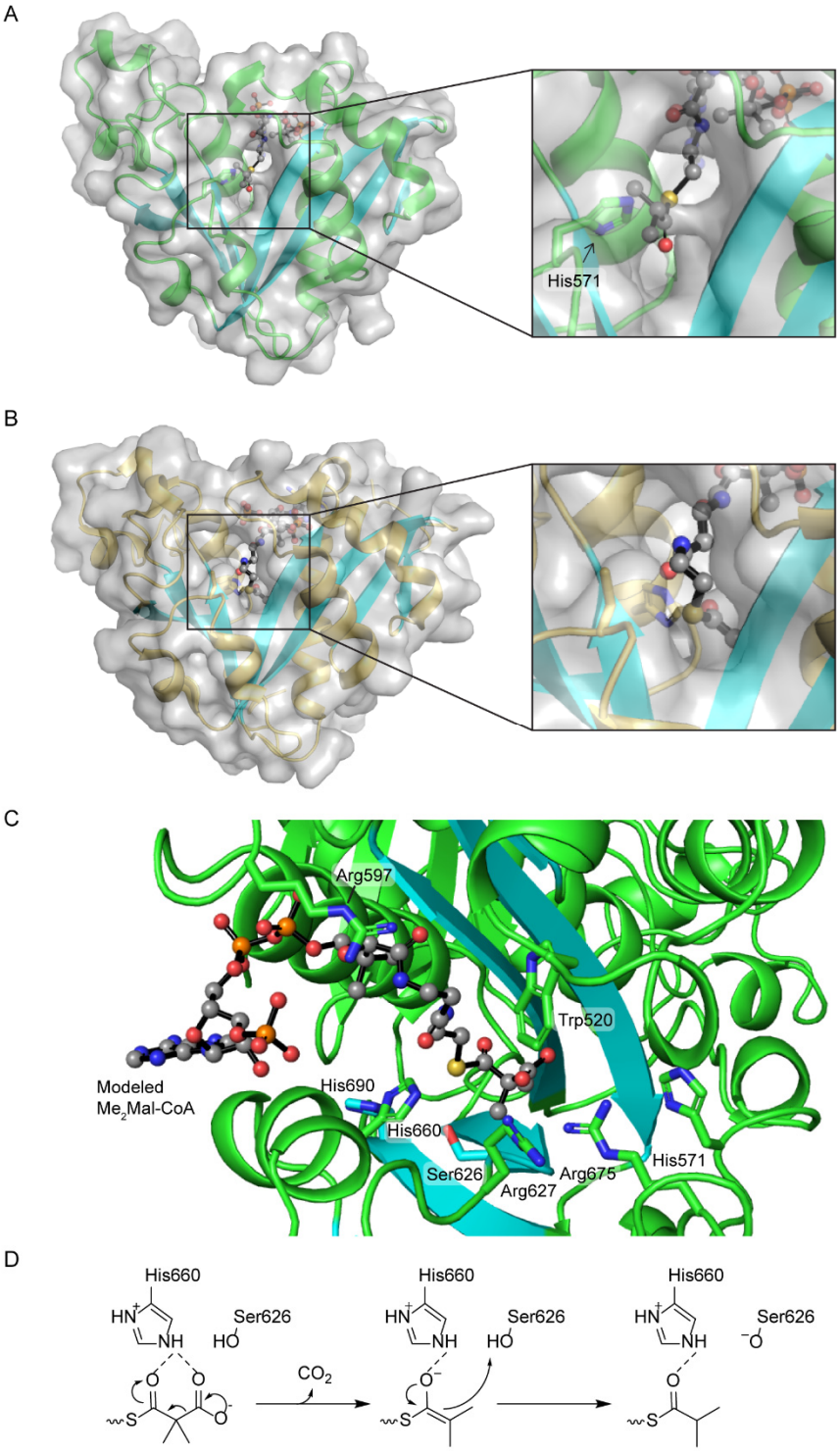


Figure S5. Distal pocket and active site of PKS GNAT-like enzymes. *Related to Figures 4 and 6.* **(A)** GphF and **(B)** CurA distal pocket. Access to the active site through the distal acyl-binding pocket (right images) is partially occluded by His571 in the GphF domain but open to solvent in CurA. **(C)** Me_2Mal modeled into GphF GNAT based on the position of isobutyryl-CoA. The terminal carboxylate can be placed near conserved Arg675. **(D)** Proposed mechanism for GphF GNAT-like decarboxylase.

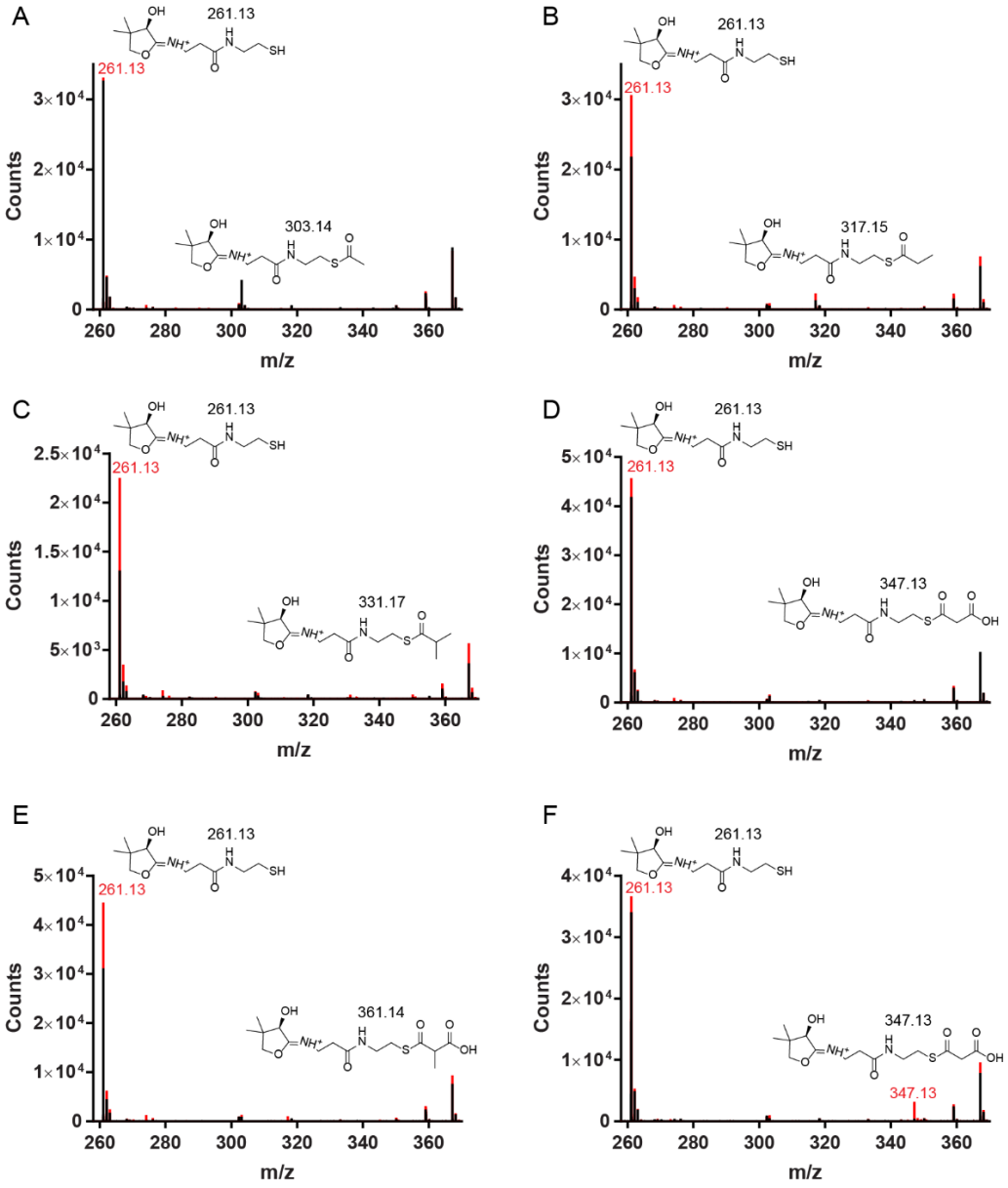


Figure S6. Representative mass spectra of GphF GNAT and MT$_L$-GNAT acyl transfer reactions. *Related to Figure 7*. The 4-hr reaction mixtures included
(A) GphF GNAT, AprA holo-ACP, acetyl-CoA;
(B) GphF GNAT, AprA holo-ACP, propionyl-CoA;
(C) GphF GNAT, AprA holo-ACP, isobutyryl-CoA;
(D) GphF GNAT, AprA holo-ACP, Mal-CoA;
(E) GphF GNAT, AprA holo-ACP, MeMal-CoA;
(F) GphF MT$_L$-GNAT; AprA holo-ACP, Mal-CoA.
AprA holo-ACP was used as a surrogate for insoluble GphF ACP$_L$. Spectra of no-enzyme controls are shown in black; reactions are shown in red. Calculated m/z values are listed in black above structures of Ppant ejection fragments, with observed m/z values in red.

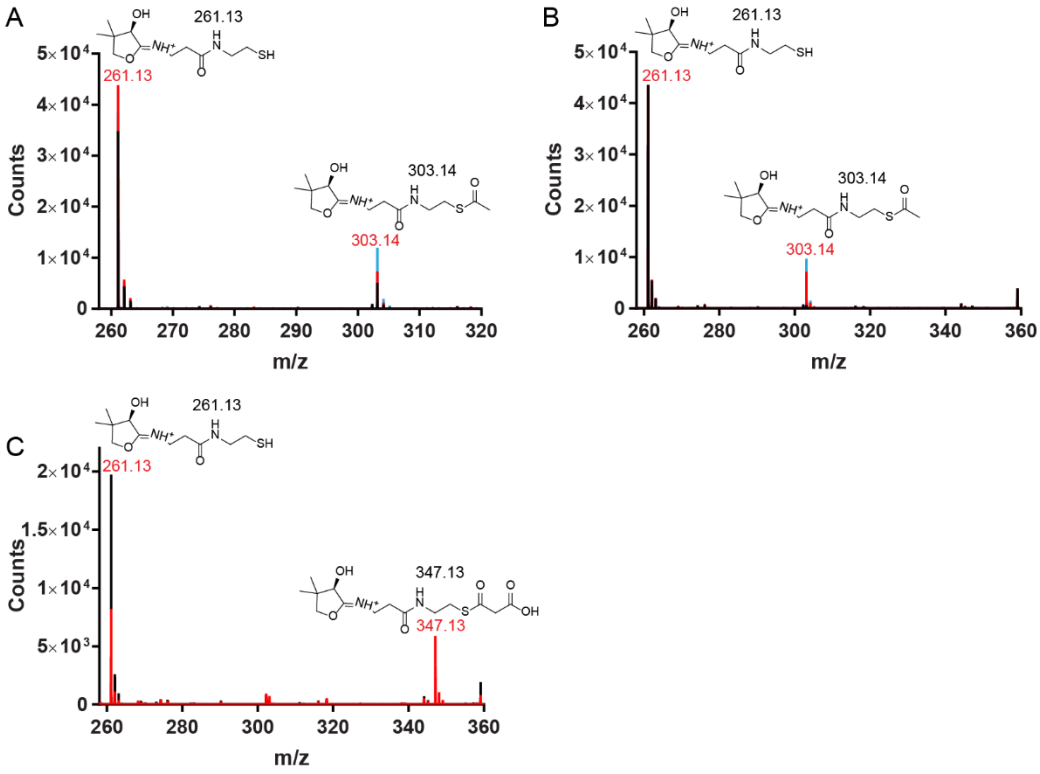


Figure S7. Representative mass spectra of CurA GNAT and *M. bouillonii* FabD acyl transfer reactions. *Related to Figure 7*. With CurA GNAT, the 4-hr reaction mixtures included
(A) CurA holo-ACP, acetyl-CoA;
(B) CurA holo-ACP, Mal-CoA.
Spectra of no-enzyme controls are shown in black, reactions with single-step-purified CurA GNAT in blue, and reactions with multistep-purified CurA GNAT in red. In (B), all Mal-ACP was decarboxylated to acetyl-ACP by CurA GNAT.
(C) Malonyl transfer reaction catalyzed by *M. bouillonii* FabD, the malonyl acyltransferase of fatty acid biosynthesis. The 5-min reaction mixtures included CurA holo-ACP and Mal-CoA. Spectra of no-enzyme controls are shown in black; reactions in red. Calculated m/z values are listed in black above structures of Ppant ejection fragments with observed m/z values in red.

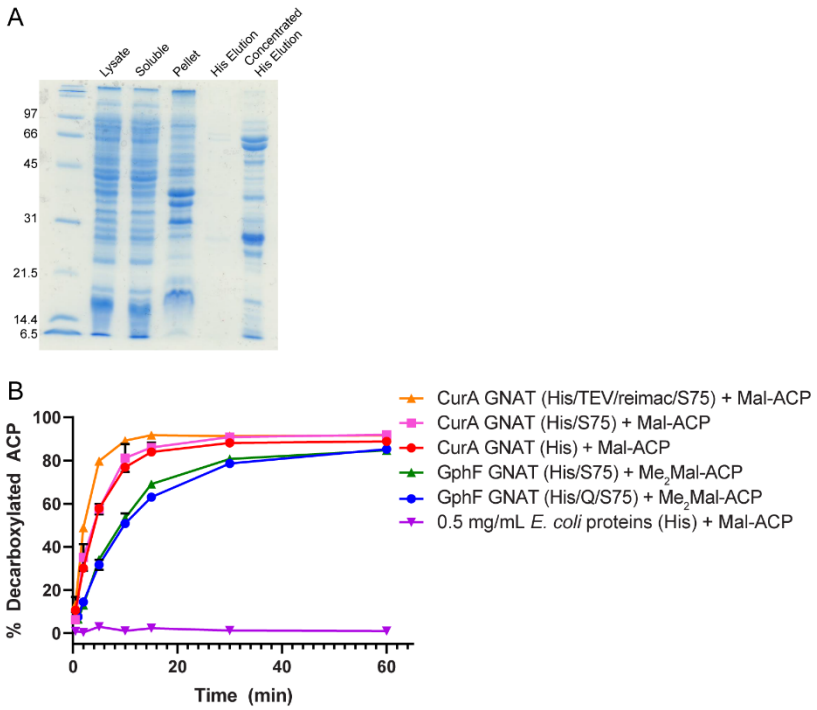


Figure S8. **Catalytic activity of GNAT-like domains in comparison to activity of an *E. coli* protein mixture enriched by Ni-NTA resin.** *Related to Figures 7 and 8*. **(A)** Mock purification to enrich non-specific *E. coli* proteins with affinity for Ni-NTA resin. The concentrated His elution was subjected to assays in (B and Figure 8). **(B)** Time courses for Mal-ACP decarboxylation by CurA GNAT subjected to the indicated purification steps, for Me_2Mal-ACP decarboxylation by GphF GNAT subjected to the indicated purification steps and for Mal-ACP by the *E. coli* protein mixture. Error bars represent triplicate experiments and, in some cases, are too small to be visible. Additional purification steps, which affect rates of acyl transfer, do not alter rapid decarboxylation activity. Non-specific *E. coli* proteins with affinity for Ni-NTA resin did not have appreciable decarboxylation activity.

Table S1. Primers. *Related to STAR Methods*.

GphF_498_F	pMAS/APS340	5'-**tacttccaatccaatgca**gcgcatcggatccgtc-3'
GphF_705_R		5'-**ttatccacttccaatgtta**gctcgcggggacc-3'
CurA_444_F	pMAS/APS442	5'-**tacttccaatccaatgca**gaagatatagatgaagttgtggctag-3'
CurA_521_R		5'-**ttatccacttccaatgtta**gcttctctctgttgtctccc-3'
GphF_H571C_F	pMAS343	5'-cggcgcctctgcaggcttcgcgggcc-3'
GphF_H571C_R		5'-ggcccgcgaagcctgcagaggcgccg-3'
GphF_R597Q_F	pMAS387	5'-gaccccgcggtccagaaccacgacttcg-3'
GphF_ R597Q _R		5'-cgaagtcgtggttctggaccgcggggtc-3'
GphF_S626A_F	pMAS426	5'-tcgtcggcgtcgcgcgttgcggc-3'
GphF_ S626A _R		5'-gccgcaacgcgcgacgccgacga-3'
GphF_S626T_F	pMAS388	5'-tcgtcggcgtcacgcgttgcggc-3'
GphF_ S626T _R		5'-gccgcaacgcgtgacgccgacga-3'
GphF_H660A_F	pMAS427	5'-ggtgctcgggttcgccctcggtcacggc-3'
GphF_ H660A _R		5'-gccgtgaccgagggcgaacccgagcacc-3'
GphF_H660N_F	pMAS438	5'-ggtgctcgggttcaacctcggtcacgg-3'
GphF_ H660N _R		5'-ccgtgaccgaggttgaacccgagcacc-3'
GphF_R675K_F	pMAS419	5'-gatcatcagcggctacaagcccgaggatgtcgac-3'
GphF_ R675K _R		5'-gtcgacatcctcgggcttgtagccgctgatgatc-3'
GphF_R675E_F	pMAS425	5'-tcatcagcggctacgagcccgaggatgtcg-3'
GphF_ R675E _R		5'-cgacatcctcgggctcgtagccgctgatga-3'
GphF_H690Y_F	pMAS390	5'-gtgtgatcgtgatctatgcgctccgtggggc-3'
GphF_ H690Y _R		5'-gccccacggagcgcatagatcacgatcacac-3'
CurA_Q326R_F	pMAS423	5'-gtaaatatcttgccagagttgagaaatcagggattgggagatcg-3'
CurA_Q326R_R		5'-cgatctcccaatccctgatttctcaactctggcaagatatttac-3'
CurA_T355S_F	pMAS384	5'-ggtgtcgaaaaagttgtggcagtatctctttgtcgcaa-3'
CurA_T355S_R		5'-ttgcgacaaagagatactgccacaacttttcgacacc-3'
CurA_L356R_F	pMAS385	5'-aaagttgtggcagtaactcgttgtcgcaattatccagac-3'
CurA_L356R_R		5'-gtctggataattgcgacaacgagttactgccacaacttt-3'
CurA_R404K_F	pMAS422	5'-tgagaaactgcttcctggttacaagcctaaagattgggaaaaccag-3'
CurA_R404K_R		5'-ctggttttcccaatctttaggcttgtaaccaggaagcagtttctca-3'
CurA_R404E_F	pMAS421	5'-tgagaaactgcttcctggttacgagcctaaagattgggaaaaccag-3'
CurA_R404E_R		5'-ctggttttcccaatctttaggctcgtaaccaggaagcagtttctca-3'
CurA_Y419H_F	pMAS405	5'-agacttgtggagttctcgtatctcacgatattcaacatcg-3'
CurA_ Y419H _R		5'-cgatgttgaatatcgtgagatacgagaactccacaagtct-3'