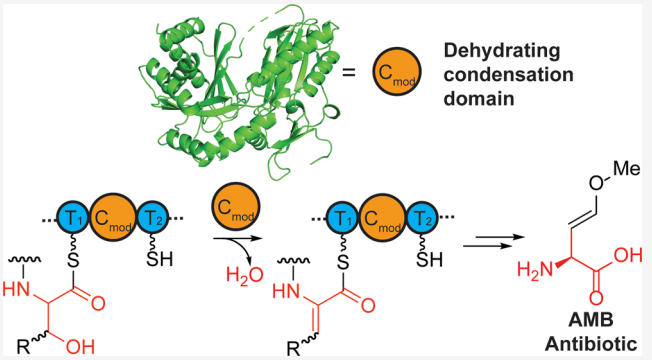
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Structure and Function of a Dehydrating Condensation Domain in Nonribosomal Peptide Biosynthesis

Jon B. Patteson,[‖] Camille Marie Fortinez,[‖] Andrew T. Putz, Juan Rodriguez-Rivas, L. Henry Bryant, III, Kamal Adhikari, Martin Weigt, T. Martin Schmeing,* and Bo Li*

Cite This: J. Am. Chem. Soc. 2022, 144, 14057−14070

ABSTRACT: Dehydroamino acids are important structural motifs and biosynthetic intermediates for natural products. Many bioactive natural products of nonribosomal origin contain dehydroamino acids; however, the biosynthesis of dehydroamino acids in most nonribosomal peptides is not well understood. Here, we provide biochemical and bioinformatic evidence in support of the role of a unique class of condensation domains in dehydration (C_{modAA}). We also obtain the crystal structure of a C_{modAA} domain, which is part of the nonribosomal peptide synthetase AmbE in the biosynthesis of the antibiotic methoxyvinylglycine. Biochemical analysis reveals that AmbE-C_{modAA} modifies a peptide substrate that is attached to the donor carrier protein. Mutational studies of AmbE-C_{modAA} identify several key residues for activity, including four residues that are mostly conserved in the C_{modAA} subfamily. Alanine mutation of these conserved residues either significantly increases or decreases AmbE activity. AmbE exhibits a dimeric conformation, which is uncommon and could enable transfer of an intermediate between different protomers. Our discovery highlights a central dehydrating function for C_{modAA} domains that unifies dehydroamino acid biosynthesis in diverse nonribosomal peptide pathways. Our work also begins to shed light on the mechanism of C_{modAA} domains. Understanding C_{modAA} domain function may facilitate identification of new natural products that contain dehydroamino acids and enable engineering of dehydroamino acids into nonribosomal peptides.



INTRODUCTION

Dehydroamino acids occur in many bioactive natural products, such as the antibiotic nisin, the anticancer drug romidepsin (Istodax), and the cyanobacterial toxin microcystin-LR (Figure S1A).[1−3] These noncanonical amino acids provide natural products with conformational rigidity, proteolytic stability, and reactive functionalities that facilitate interactions with their biological targets.[4] For example, the α,β-dehydroalanine (Dha) in microcystin-LR forms a covalent linkage with a nucleophic cysteine in the target.[5] In addition, the unique electronic properties and chemical reactivities of α,β-dehydroamino acids enable them to participate in diverse chemical transformations, including nucleophilic-, radical-, and cyclo-additions in biomolecules under mild conditions.[6] Biosynthesis of dehydroamino acids has been extensively studied in ribosomally synthesized and posttranslationally modified peptide (RiPP) natural products, such as nisin, whereby Dha and dehydrobutyrine (Dhb) are installed by lanthipeptide dehydratases via dehydration of serine and threonine, respectively.[7] Dha/Dhb can subsequently be converted to lanthionines, lysinoalanines, pyridines, and D-amino acids in RiPP biosynthesis, which highlights the versatility of dehydroamino acid intermediates in biosynthetic transformations.[8]

Many nonribosomal peptides contain dehydroamino acids, such as the aforementioned romidepsin and microcystin-LR (Figure S1A),[2,3] or are hypothesized to be derived from dehydroamino acid intermediates, such as the antitumor drug bleomycin and azabicyclene (Figure S1B).[9,10] Nonribosomal peptides are a major class of natural products that exhibit wide-ranging biological activities and therapeutic applications. The biosynthesis of nonribosomal peptides requires nonribosomal peptide synthetases (NRPSs), assembly lines made up of repeating sets of domains that organize into modules.[11,12] A typical module contains an adenylation (A) domain that activates an amino acid, a thiolation (T) domain where the amino acid is tethered via a thioester, and a condensation (C) domain that forms an amide bond between the amino acids on the donor and acceptor T domains. Once peptide elongation completes, a terminal thioesterase (TE) domain cleaves the

Received: December 20, 2021
Published: July 27, 2022

https://doi.org/10.1021/jacs.1c13404
J. Am. Chem. Soc. 2022, 144, 14057−14070

thioester, which releases the free peptide from the T domain. Using this assembly line strategy with assistance from tailoring enzymes, NRPSs bypass the ribosome and incorporate nonproteinogenic amino acids into nonribosomal peptides at different stages of biosynthesis, including directly activating the nonproteinogenic amino acid using the A domain, modifying the amino acid after its loading onto a T domain, or modifying the amino acid on a peptide product. The enamine of dehydroamino acids is unstable and prone to hydrolysis; therefore, the biosynthesis of Dha or Dhb in bleomycin, mycrocystin-LR, and syringomycin was proposed in the 2000s to involve dehydration of serine or threonine tethered to a T domain on the NRPS assembly line.[3,9,13,14] However, the timing of dehydration was unclear (*i.e.*, whether dehydration occurs on a T domain-tethered amino acid or a tethered peptide). It was also unknown whether an NRPS domain or an auxiliary tailoring enzyme is responsible for dehydration.

A phylogenetic study of C domains in 2012 showed that two C domains of the bleomycin NRPSs and a C domain of the microcystin-LR NRPS form a distinct clade from the other C domains.[15] These C domains were proposed to participate in modifying the amino acid incorporated on the NRPS assembly line and were assigned the name "modified AA" (modAA). We will refer to these C domains as C_{modAA}. Our expanded bioinformatic analysis found that C_{modAA} domains also exist in NRPSs involved in the biosynthesis of other nonribosomal peptides that contain dehydroamino acids or are likely derived from dehydroamino acid intermediates.[16] Despite the compelling bioinformatic evidence for a role of the C_{modAA} domain in dehydration, no experimental evidence existed to support this function until very recently in the nonribosomal peptide albopeptide, which contains two consecutive Dha and Dhb.[17] *In vitro* reconstitution of the NRPS, AlbB, in albopeptide biosynthesis confirmed the function of two AlbB-C_{modAA} domains in the dehydration of Ser and Thr to generate Dha and Dhb, respectively. Although this study provided important biochemical evidence for the function of C_{modAA} domains, no structural information on C_{modAA} domains was available and mechanistic understanding of C_{modAA} domains was also lacking.

Notably, a C domain in nocardicin biosynthesis, NocB-C_5, also generates a dehydroamino acid intermediate and uses this intermediate to synthesize the β-lactam ring in nocardicin.[18,19] NocB-C_5 appears to be a unique member of the $^{D}C_{L}$ subfamily of C domains that typically catalyze condensation between a D- and an L-amino acid. The $^{D}C_{L}$ subfamily is distinct from C_{modAA} domains, which is a different way to introduce dehydroamino acids by NRPSs. While detailed mutational analysis of NocB-C_5 has identified key residues for activity, the divergence in the sequence of NocB-C_5 from C_{modAA} domains (Figure S2A) suggests that different residues may be required for the activity of C_{modAA} domains.

We identified a C_{modAA} domain in the biosynthesis of methoxyvinylglycine or L-2-amino-4-methoxy-*trans*-3-butenoic acid (AMB, Figure 1A), an antimicrobial nonproteinogenic amino acid produced by *Pseudomonas aeruginosa* that arrests plant seed germination.[16,20] AMB is produced as an alanyl-AMB dipeptide (Ala-AMB) by an NRPS pathway, which requires two NRPSs, AmbB and AmbE, and two nonheme iron, α-ketoglutarate-dependent oxygenases, AmbC and AmbD.[16] AmbB contains a typical domain order of A-T-C, which activates and loads alanine prior to using alanine as the donor in condensation. In contrast, AmbE's composition of Q-A-MT-T_1-C_{modAA}-T_2-TE is unusual in several aspects: it

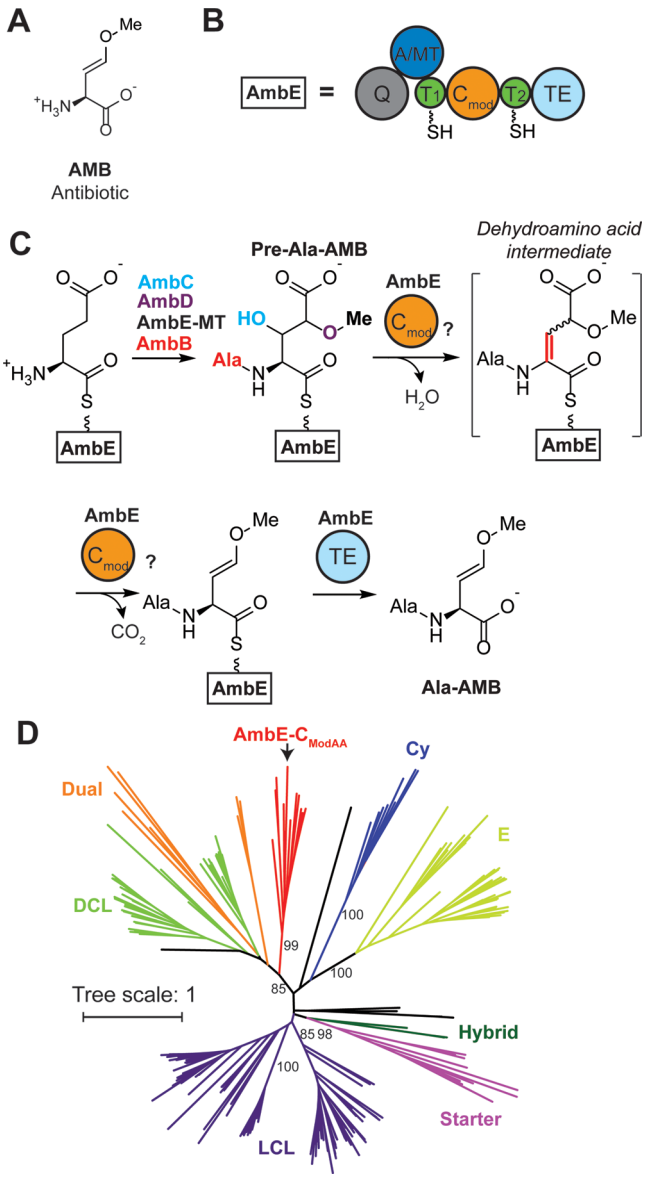


Figure 1. Identification of a condensation (C) domain that may catalyze dehydration in the synthesis of nonribosomal peptides. (A) Structure of L-2-amino-4-methoxy-*trans*-3-butenoic acid (AMB). (B) Domain organization of AmbE in colored spheres (A: adenylation, C_{modAA}: condensation with a role in modifying amino acids, MT: methyltransferase, T: thiolation, TE: thioesterase, Q: unknown function). The wavey line and SH indicate the phosphopantetheine (ppant) arm of T domains. (C) Biosynthesis of Ala-AMB. A C_{modAA} domain (C_{mod} for short) is proposed to catalyze dehydration of pre-Ala-AMB to Ala-AMB that is tethered to the NRPS AmbE. (D) Unrooted maximum likelihood phylogenetic tree of 199 C domains including three modAA domains from the NaPDoS database and nine other C_{modAA} domains from known natural product pathways including AmbE-C_{modAA}. These C_{modAA} domains (red) form a unique subfamily of C domains. The tree scale represents the average expected percentage (1 = 100%) of amino acid substitutions per site. Bootstrap values greater than 80 are labeled next to the branches. Abbreviations: LCL, condensation between two L-amino acids (purple); DCL, condensation between a D- and an L-amino acid (green); dual, condensation and epimerization (orange); starter, acylation to various molecules (pink); hybrid, condensation of amino acids to polyketides (dark green); Cy, condensation and heterocyclization (blue); and E, epimerization (yellow). The sequences in black do not belong to any of these groups. The modAA clade

https://doi.org/10.1021/jacs.1c13404
J. Am. Chem. Soc. 2022, 144, 14057−14070

Figure 1. continued

contains C_{modAA} domains from the following sequences: AmbE (AAG05690.1), AzeB (AAG06715.1, azabicyclene biosynthesis), DepE (ABP57749.1, romedepsin biosynthesis), BlmX (AAG02359.1, bleomycin biosynthesis), BlmVI (AAG02355.1, bleomycin biosynthesis), LgnD (AIZ66879.1, legonmycin biosynthesis), HasO (CZT62784.1, hassallidin biosynthesis), Zmn17 (CCM44337.1, zeamine biosynthesis), NdaA (ATP76243.1, nodularin biosynthesis), McyA (BAA83992.1, microcystin biosynthesis), PuwF (AIW82283.1, puwainaphycin biosynthesis), and PuwG (AIW82284.1, puwainaphycin biosynthesis).

contains a domain of the unknown function (Q domain) at the N terminus, and there is no A domain between the C_{modAA} and T_2 in the final module, so two T domains directly flank C_{modAA} (Figure 1B). The Q domain shares low sequence identities with X domains in vancomycin biosynthesis and interface (I) domains in siderophore biosynthesis (Figure S2B). AmbE activates and loads glutamate on T_1 followed by sequential hydroxylation by AmbC and AmbD, which produces a β,γ-dihydroxyglutamate intermediate linked to AmbE (Figure 1C).[16] Subsequently, a methyltransferase (MT) domain in AmbE catalyzes methylation of the γ-hydroxyl group and the C domain in AmbB condenses the modified glutamate with an alanine, which yields Ala-β-hydroxy-γ-methoxy-Glu (hereafter referred to as pre-Ala-AMB) (Figure 1C).[16] Characterization using deuterium-labeled glutamate as the substrate revealed that the α-proton/deuterium is removed during the conversion from pre-Ala-AMB to Ala-AMB, which suggests that pre-Ala-AMB undergoes α,β-dehydration to a cryptic dehydroamino acid intermediate that enables decarboxylation to generate Ala-AMB (Figure 1C).[16] AmbE-C_{modAA} belongs to the modAA subfamily of C domains (Figure 1D), and we proposed that AmbE-C_{modAA} catalyzes the cryptic α,β-dehydration in the last steps of Ala-AMB biosynthesis.[16]

Here, we characterize the structure and function of the C_{modAA} domain in AMB biosynthesis and provide experimental support for its role as a dehydrating condensation domain. We identified important residues of AmbE-C_{modAA} for dehydrative decarboxylation by mutagenesis and studied the mutants under both multiple- and single-turnover conditions. We bioinformatically analyzed 27 C_{modAA} domains in biosynthetic pathways for known natural products and suggest that these C_{modAA} domains are responsible for incorporating α,β-dehydroamino acids in these natural products that enable diverse biosynthetic transformations.

RESULTS

AmbE-C_{modAA} Transforms pre-Ala-AMB to Ala-AMB on the Donor T_1 Domain. To characterize the function of the C_{modAA} domain in AMB biosynthesis and determine if C_{modAA} acts on its donor T (T_1) or acceptor T domain (T_2), we generated AmbE constructs containing different domain combinations. The TE domain was omitted to prevent the release of the product from the T domains of the NRPSs. A one-pot reaction was performed including AmbB, AmbC, AmbD (AmbBCD hereafter), and different AmbE constructs in the presence of all necessary substrates and cofactors. Cysteamine was used in these one-pot reactions to capture the intermediates that accumulate on the NRPS and to probe thiotemplated biosynthesis (Figure S3).[16,21] The captured cysteamine adducts were characterized using liquid chromatography-coupled high-resolution mass spectrometry (LC-HRMS) (Figure 2). We expressed truncates of AmbE that ended before C_{modAA} (Q-A-MT-T_1) or after C_{modAA} (Q-A-MT-T_1-C_{modAA}). The one-pot reaction containing AmbE-Q-A-MT-T_1 and AmbBCD yielded captured pre-Ala-AMB (Figure S4B). Formation of pre-Ala-AMB without C_{modAA} indicates that pre-Ala-AMB accumulates on T_1 prior to the action of C_{modAA} (Figure 2). When Q-A-MT-T_1-C_{modAA} was used in the reaction, captured Ala-AMB was detected (Figures 2B, S4, and S5). Results from these reactions suggest that C_{modAA} transforms pre-Ala-AMB to Ala-AMB on the donor T_1 domain.

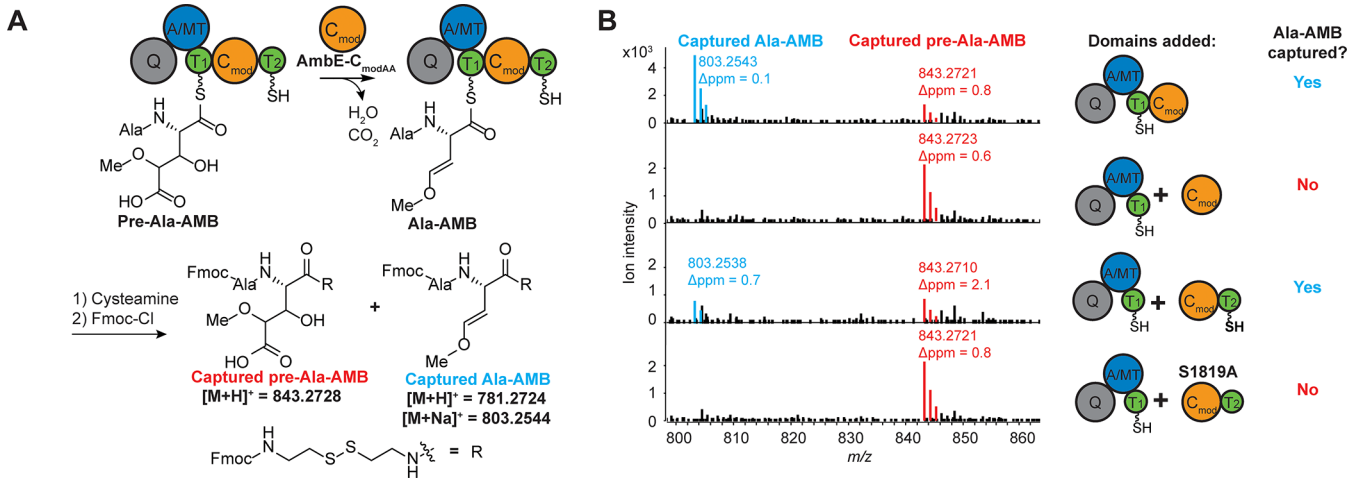


Figure 2. Transformation of pre-Ala-AMB to Ala-AMB requires C_{modAA} in *cis* and C_{modAA}-T_2 in *trans*. (A) One-pot reaction scheme. When different combinations of AmbE constructs are used instead of full-length AmbE, late-stage biosynthetic intermediates accumulate on the T domains and are released and captured by cysteamine and detected by LC-HRMS. (B) LC-HRMS analysis of captured pre-Ala-AMB or Ala-AMB (structures shown in A). A wider mass window that includes both $[M + H]^+$ and $[M + Na]^+$ ions for each compound is shown in Figure S4A. Δppm, parts per million error between the calculated and observed mass-to-charge ratios (m/z). Additional controls are shown in Figure S5. Ala-AMB formation requires attachment of C_{modAA} to T_1 or *holo* T_2 in the split system. S1819A indicates an alanine mutation of the conserved serine in T_2 where the ppant arm is attached.

https://doi.org/10.1021/jacs.1c13404
J. Am. Chem. Soc. 2022, 144, 14057−14070

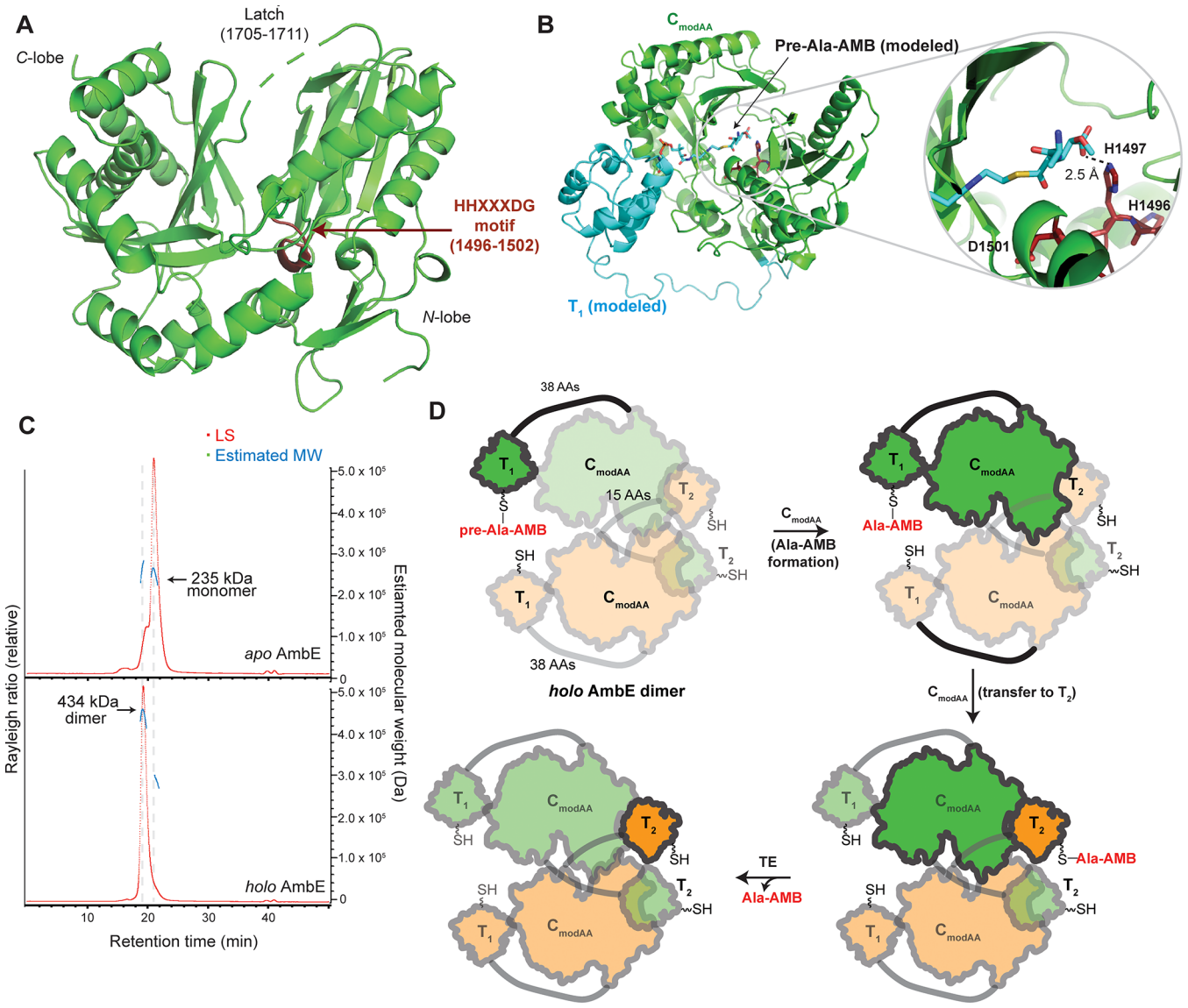


Figure 3. Crystal structure of AmbE-C_{modAA} and oligomeric states of full-length AmbE. (A) Overall structure of AmbE-C_{modAA} (PDB: 7R9X) with the active site tunnel between the N- and C-terminal lobes. The conserved HHXXXDG motif is located in the active site tunnel (Figure S9B). (B) Modeling of pre-Ala-AMB-bound T_1 at the canonical donor binding site of C_{modAA} positions pre-Ala-AMB in proximity to the HHXXXDG motif in the active site tunnel. (C) SEC-MALS analysis shows that *apo* AmbE is a monomer, while *holo* AmbE is a dimer. MW, molecular weight. LS, light scattering. (D) Model for the function of dimeric *holo* AmbE. The C_{modAA} of one AmbE protomer (green) interacts with the T_2 of the other protomer (orange) as the acceptor T domain.

We also tested the activity of C_{modAA} in *trans* by separating AmbE into Q-A-MT-T_1 and C_{modAA}-T_2. Incubation of C_{modAA}-T_2 with AmbBCD and Q-A-MT-T_1 in the one-pot reaction with cysteamine yields captured Ala-AMB, albeit at a lower level than Q-A-MT-T_1-C_{modAA} (Figures 2B, S4, and S5). This result further supports the requirement of C_{modAA} for the biosynthesis of Ala-AMB and demonstrates that the split system is functional for investigating C_{modAA} activity. Truncation of C_{modAA}-T_2 to only include the C_{modAA} domain in the reaction still resulted in similar levels of captured pre-Ala-AMB but no captured Ala-AMB (Figure 2B). This surprising result suggests that T_2 is required for Ala-AMB production in the split system when C_{modAA} is separated from Q-A-MT-T_1. Additionally, mutation of the catalytic S1819 residue on T_2 to Ala in C_{modAA}-T_2 abolishes Ala-AMB formation. This mutation prevents the installment of the phosphopantetheine (ppant) arm on T_2. The lack of Ala-AMB formation by this mutant indicates that *holo* T_2 is needed for the split system to function (Figures 2B and S5). Together, these experiments reveal that pre-Ala-AMB is converted to Ala-AMB by C_{modAA} while attached to T_1 and that C_{modAA} must be part of a construct including the upstream module (Q-A-MT-T_1) or downstream domain (*holo* T_2) to be functional.

Because *holo* T_2 is required for the split system to function, we further examined the possibility that C_{modAA} might convert pre-Ala-AMB to Ala-AMB on T_2 instead of T_1. We previously found that the full-length AmbE T_2 mutant (S1819A) did not release free Ala-AMB in the one-pot reaction.[16] Adding cysteamine to this reaction only resulted in captured pre-Ala-AMB.[16] Although these results seemed to suggest that formation of Ala-AMB might take place on T_2, reanalysis of the cysteamine capture data revealed that the S1819A mutant produces free Ala-AMB that is not linked to cysteamine (Figure S6). This unexpected result shows that the T_2 mutant

https://doi.org/10.1021/jacs.1c13404
J. Am. Chem. Soc. 2022, 144, 14057−14070

is able to convert pre-Ala-AMB to Ala-AMB and supports the conclusion that C_{modAA} modifies pre-Ala-AMB on T_1.

Production of free Ala-AMB by the T_2 mutant may result from the action of the AmbE TE domain, which could cleave cysteamine from the Ala-AMB-cysteamine adduct. Consistent with this proposal, adding cysteamine to the one-pot reaction of the AmbE TE mutant (S1956A) generated abundant cysteamine-captured Ala-AMB but little free Ala-AMB (Figure S6). Furthermore, adding cysteamine to the wild-type AmbE reaction also boosted the production of free Ala-AMB (Figure S6). Together, these data support a model in which *holo* T_2 is required for the transfer of Ala-AMB and subsequent cleavage by TE in full-length AmbE. In the split system, *holo* T_2 may help stabilize C_{modAA} or facilitate interactions between C_{modAA} and T_1.

AmbE-C_{modAA} Transfers Ala-AMB to the Acceptor T_2 Domain. The split AmbE system was analyzed using size-exclusion chromatography to determine if Ala-AMB is transferred to the acceptor T_2 domain. Upon completion of a scaled-up one-pot reaction for Ala-AMB biosynthesis, Q-A-MT-T_1 and C_{modAA}-T_2 were separated by size-exclusion chromatography, and the intermediates linked to Q-A-MT-T_1 and C_{modAA}-T_2 were captured by cysteamine (Figures S7, S8). Pre-Ala-AMB was only detected on Q-A-MT-T_1, while Ala-AMB was detected on both Q-A-MT-T_1 and C_{modAA}-T_2 (Figure S7). This result supports that Ala-AMB is formed on T_1 and subsequently transferred to T_2. Detection of Ala-AMB on T_1 suggests that the transfer of Ala-AMB from T_1 to T_2 is less efficient in the split system than in full-length AmbE.

Crystallography of AmbE-C_{modAA}. We conducted structural studies of AmbE-C_{modAA} to further understand the activity of this domain. Purified C_{modAA} was subject to crystallization screening and initial conditions were optimized to yield crystals suitable for diffraction. We obtained a 2.1 Å resolution structure of AmbE-C_{modAA} with a R_{free} of 23% (Table S2). Two copies of C_{modAA} exist in the asymmetric unit in near-identical conformation. AmbE-C_{modAA} possesses an overall structure similar to other proteins in the C domain family:[22] a V-shaped pseudodimer consisting of N- and C-terminal lobes with a chloramphenicol acetyltransferase (CAT) fold (Figure 3A). These lobes are known to exist in different relative orientations or "openness" in previously determined, catalytically active C domains, and AmbE-C_{modAA} falls in the middle of this observed range (Figure S9A).[23,24] The junction of the lobes forms the classic active site tunnel that connects the canonical donor and acceptor T domain-binding sites with the active site (Figure S9B), where the typical C domain HHXXXDG motif (H1496−G1502) is located (Figures 3A and S9B). Notably, the latch element (H1705−P1711) above the active site is disordered in both copies of C_{modAA}, which suggests flexibility in this region (Figure 3A).

We combined the structure of C_{modAA} with coevolutionary analysis and modeling to investigate potential interactions with its donor and acceptor T domains. For coevolutionary analysis of T_1 and C_{modAA}, we extracted 1534 T/C_{modAA} pairs in which T and C_{modAA} are present within the same protein and separated by 100 or fewer residues. These T/C_{modAA} pairs were aligned and subject to filtered direct coupling analysis (filterDCA).[25] Evolutionarily conserved contacts identified from DCA were mapped on a homology model of T_1 that is positioned at the canonical donor T domain-binding site of C_{modAA}.[26] The distances observed between the DCA-predicted T/C_{modAA} pairs support this canonical binding mode between T_1 and C_{modAA} (Figure S9C and Table S3). Thus, we used existing T-C structures with *holo* T domains as a guide to build a model of pre-Ala-AMB-T_1 bound to C_{modAA} (Figure 3B). The model shows that pre-Ala-AMB can be positioned at the active site within the hydrogen bonding distance of residue H1497 (HHXXXDG motif, Figures 3B and S9D). Since donor T domains coevolve with C domains,[18] our coevolutionary analyses indicate that T_1 likely binds C_{modAA} at the position observed for regular and terminal C domains.[26]

In contrast, DCA between C_{modAA} and the acceptor T_2 failed to give a strong signal. Furthermore, when a homology model of T_2 is positioned at the acceptor binding site analogous to that previously observed in C domain crystal structures,[26−29] the N terminus of T_2 is too far away from the C terminus of C_{modAA}: the ∼67 Å distance cannot be bridged by the 15-residue linker that exists between C_{modAA} and T_2 in AmbE (Figure S10A, left). We attempted to model alternative binding modes of T_2 to C_{modAA} where the N terminus of T_2 is closer to the C terminus of C_{modAA} for the short C_{modAA}-T_2 linker to bridge the distance (Figure S10A, middle and right), but there is no position that allows the pantetheine arm to be placed into the canonical acceptor site tunnel. AmbE-C_{modAA} contains an auxiliary opening to the active site (Figure S10B). T_2 could bind here between the two central sheets of the CAT folds of each lobe around residues 1582 of the N lobe and 1722 of the C lobe (Figure S10B). This binding mode would be permitted by the 15-residue linker between C_{modAA} and T_2; however, it has not been observed in NRPS biology, and an analogous opening is present in some C domains that are thought to have canonical acceptors (Figure S10C),[30−32] and in the X domain, a noncatalytic C domain family member.[33]

Alternatively, the C_{modAA}:T_2 linker does not need to bridge the distance between the C terminus of C_{modAA} and the N terminus of T_2. If AmbE exists as a homooligomer, T_2 from one protomer of AmbE could interact with C_{modAA} from the other. Size-exclusion chromatography reveals that purified AmbE contains two species that correspond to a monomer and dimer, with the monomer being the major species. Size-exclusion chromatography-coupled multiangle light scattering (SEC-MALS) analysis of the isolated monomer peak in reducing buffer supports the assignment of the monomer (Figure 3C). Incubating the monomer with the phosphopantetheinyl transferase Sfp in the presence of coenzyme A and $MgCl_2$ converts the monomer to the dimer (Figures 3C and S11), suggesting that phosphopantetheinylation by Sfp switches the oligomeric state of AmbE. Indeed, the use of the monomeric or dimeric AmbE in a one-pot assay shows Ala-AMB production by the dimer but not the monomer (Figure S12). The activity of the dimeric species and inactivity of the monomeric species is consistent with the dimer being *holo* and the monomer being *apo*. C_{modAA} is monomeric in solution and in the crystallized structure, thus, the architecture of the AmbE dimer is unknown. Nonetheless, dimerization of AmbE could permit C_{modAA} to access the canonical acceptor tunnel in *trans* in spite of the short linker between C_{modAA} and T_2. To further probe which domains of AmbE are required for dimerization, we also examined the oligomeric states of C_{modAA}-T_2 and Q-A-MT-T_1-C_{modAA}. Both constructs exist as monomers regardless of being *apo* or *holo* (Figures S13 and S14A). Furthermore, the S1819A mutant of full-length AmbE, which contains a mutation in T_2 that prevents phosphopantetheinylation, only exists as a monomer regardless of the *apo* or *holo* state of T_1 (Figure S14B). Collectively, these results suggest that AmbE

https://doi.org/10.1021/jacs.1c13404
J. Am. Chem. Soc. 2022, 144, 14057−14070

dimerization requires *holo* T_2 as well as the domains before C_{modAA}-T_2.

Based on the conversion of pre-Ala-AMB to Ala-AMB on T_1, the transfer of Ala-AMB to T_2, and the dimeric state of *holo* AmbE, we propose the following steps in late-stage biosynthesis of Ala-AMB (Figure 3D): AmbE-C_{modAA} catalyzes conversion of pre-Ala-AMB to Ala-AMB on T_1, followed by the transfer of Ala-AMB from T_1 of one AmbE protomer to T_2 of the other protomer; then AmbE TE catalyzes the hydrolytic release of Ala-AMB from T_2 into solution.

Mutagenesis of the C_{modAA} Domain. We sought to identify important residues for C_{modAA} function by structural modeling, sequence alignment, and mutagenesis. Modeling of C_{modAA} bound to pre-Ala-AMB-T_1 and the ppant arm of *holo* T_2 positions pre-Ala-AMB in the vicinity of residues H1496, H1497, and D1501 of the HHXXXDG motif[22,34,35] (Figures 3B and S9D). Sequence alignment of 405 nonidentical C_{modAA} domains highlights residues that are mostly conserved within C_{modAA} domains (Figures S15, S16), including R1605, H1632, T1650, and N1700, which are located near the active site (Figures 4A, and S17). Several charged residues of AmbE-C_{modAA} that are not conserved in other C_{modAA} domains are also positioned in the active site tunnel, such as D1726, D1728, D1734, and E1736 (Figures 4A and S17). We made a total of 20 single mutations to identify residues that are important for AmbE-C_{modAA} activity. These mutations were introduced in full-length AmbE to assess how they affected the ability of AmbE to synthesize and release Ala-AMB (Figure 4B). The D1696A and N1700A mutants of AmbE produce Ala-AMB at a higher level than that of wild-type (WT). In contrast, the H1632A and D1726A mutants exhibit lower Ala-AMB production than the WT (Figure 4B). Notably, single mutations in the HHXXXDG motif to Ala (H1496A, H1497A, and D1501A) do not alter Ala-AMB production significantly. However, the H1497A/D1501A double mutation abolishes Ala-AMB production (Figure 4B). LC-MS analysis shows that the AmbE mutants produce a single species of Ala-AMB with the same retention time as the product of the AmbE WT. This observation suggests that all products are in the expected L-Ala-L-AMB configuration and that these mutations do not alter stereochemistry.

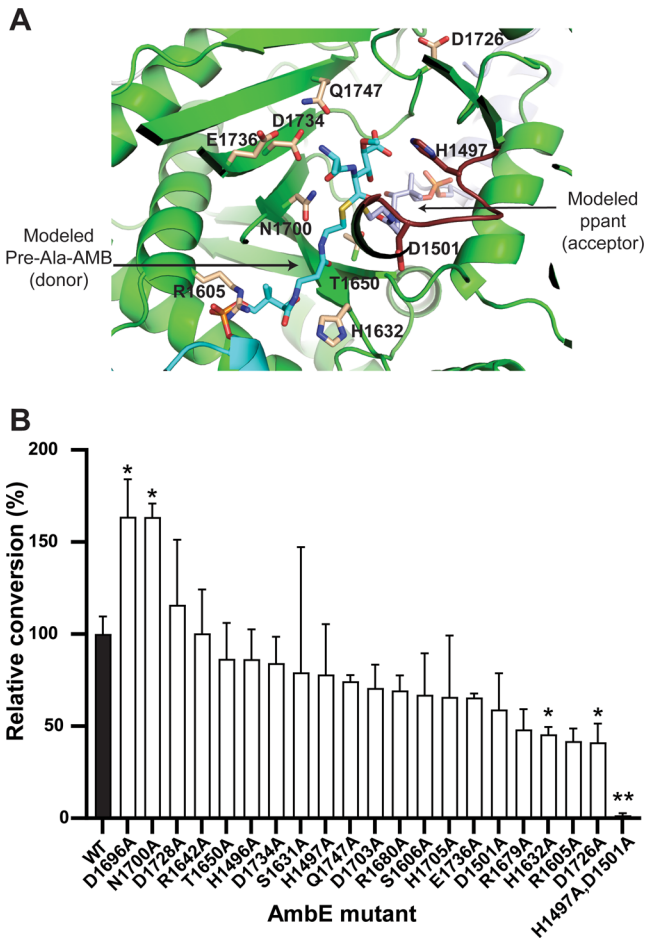


Figure 4. Mutagenesis of AmbE-C_{modAA}. (A) Active site of AmbE-C_{modAA} with modeled donor pre-Ala-AMB-ppant and acceptor ppant. (B) Relative % conversion of Ala to Ala-AMB by full-length AmbE WT and mutants. The bar graph depicts the average and standard error of at least two independent replicates for each protein. Statistical tests were performed for each mutant against WT (see the Methods section): $*p < 0.05$, $**p < 0.01$.

AMB biosynthesis involves at least 10 transformations;[16] we reasoned that the reaction catalyzed by AmbE-C_{modAA} may not be the rate-limiting step and therefore reduction in AmbE-C_{modAA} activity caused by single mutations might not be detected in the one-pot reconstitution assay. Thus, we used the Q-A-MT-T_1 and C_{modAA}-T_2 split system to first accumulate pre-Ala-AMB on Q-A-MT-T_1 and then add C_{modAA}-T_2 to measure the kinetics of C_{modAA}-T_2 with the goal of directly comparing the rates between WT C_{modAA}-T_2 and the mutants (Figure 5A). We selected 11 residues of C_{modAA} that are either projected into or located near the active site tunnel for mutation in C_{modAA}-T_2. Three C_{modAA}-T_2 mutants, H1496A, D1501A, and D1728A, were insoluble. We purified the other eight C_{modAA}-T_2 mutants and confirmed that they contain similar secondary structures as the WT by circular dichroism (Figure S18). The relative rates of conversion from pre-Ala-AMB to Ala-AMB catalyzed by WT and mutants of C_{modAA}-T_2 were obtained (Figures S19 and S20). The T1650A and N1700A mutants of C_{modAA}-T_2 exhibit increased rates of Ala-AMB formation compared to WT, whereas E1736A shows a similar rate as that of WT (Figures 5B and S20). The Q1747A, D1726A, and R1605A mutations significantly lower the rate of Ala-AMB formation to 20, 17, and 4% of WT, respectively. The H1497A and D1734A mutations abolish Ala-AMB production (Figures 5B and S20).

Bioinformatic Analysis of C_{modAA} Domains. To expand the understanding of C_{modAA} functions in different natural product pathways, we compiled 13 natural products whose biosynthetic gene clusters encode C_{modAA} domains from the MIBiG database (latest version, October 2019).[36] We also curated additional seven natural products whose gene clusters contain C_{modAA} domains (Table S4). We proposed the function of each C_{modAA} domain by correlating the amino acid that we predict based on the specificity of the A domain preceding C_{modAA}, with the natural product structure, and then cross-referenced the proposal with the literature.[9,10,17,37] Of the 27 C_{modAA} domains from 20 biosynthetic pathways, 14 C_{modAA} domains from 10 pathways replace a regular C domain and correlate with the incorporation of an upstream Ser or Thr in the precursor and Dha or Dhb in the natural product, respectively, such as microcystin and romidepsin (Figures S1 and 6A,B).[2,3] This correlation supports a dual role of C_{modAA} domains in dehydrating Ser/Thr to Dha/Dhb and forming the amide bond during peptide elongation. The remaining 10

https://doi.org/10.1021/jacs.1c13404
J. Am. Chem. Soc. 2022, 144, 14057−14070

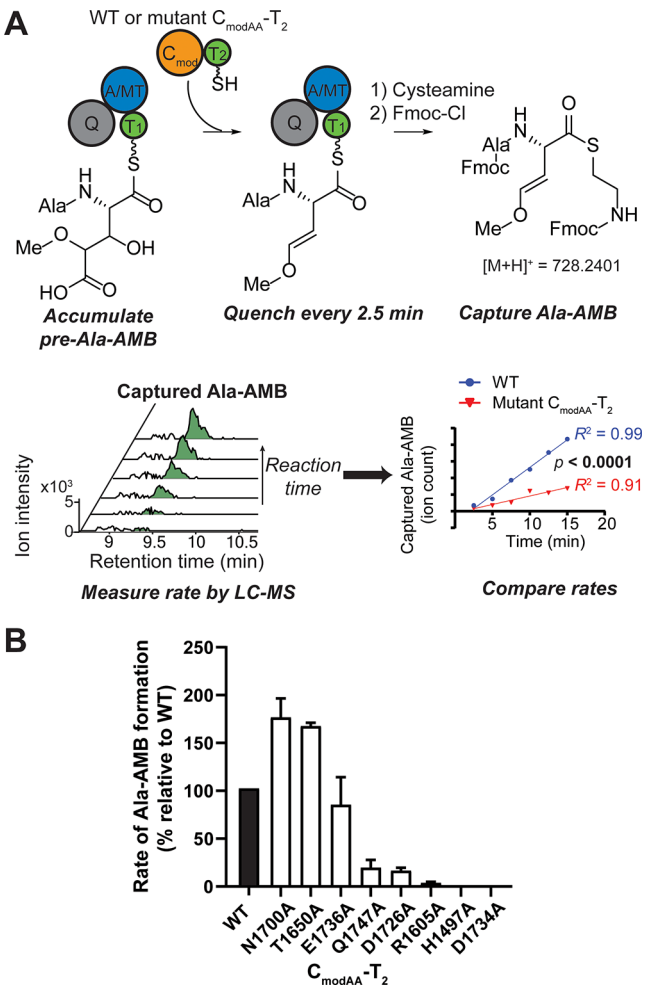


Figure 5. Rate analysis of AmbE-C$_{modAA}$-T$_2$ mutants. (A) Determination of the relative rates of Ala-AMB formation catalyzed by C$_{modAA}$-T$_2$ mutants. Representative graphs comparing the activity of WT C$_{modAA}$-T$_2$ and the Q1747A mutant by measuring Ala-AMB formation over time. R^2 values for the linear fit are shown for each graph. (B) Normalized reaction rates of C$_{modAA}$-T$_2$ mutants relative to that of WT. Three independent experiments were performed for each mutant except for R1605A, which was tested twice, and error bars depict standard errors of the mean from two or three experiments. Because in every experiment the rate of the WT control is set to 100% for calculation of the relative rate of each mutant, the standard error of the mean for the WT rates is zero. Raw data from each independent experiment are shown in Figure S20 and Table S5. The reaction rates of all mutants except for E1736A are significantly different from that of WT. For each mutant, statistical significance (p) was calculated using analysis of covariance to compare the slopes of WT and the mutant in the same experiment (Figure S20).

natural products including AMB do not contain dehydroamino acids (Figure S21). Analysis of their biosynthetic pathways suggests that dehydroamino acid intermediates are further modified via various transformations, including conjugate addition with an amine in prezeamine, pyrrolizidine formation in azabicyclene, and conjugate addition and pyrimidine formation in bleomycin (Figures 6B and S21).[9,10,38] The C$_{modAA}$-catalyzed dehydration is a central step and unifying theme for these pathways (Figure 6A).

Phylogenetic analysis of the 27 C$_{modAA}$ domains reveals that some C$_{modAA}$ domains cluster by function (Figures 6B and S22). AmbE-C$_{modAA}$ does not group well with other C$_{modAA}$ domains, which is consistent with its unique dehydrating and decarboxylating function in addition to catalyzing trans-thioesterification rather than amide bond formation. Nine C$_{modAA}$ domains with additional functions post dehydration form three distinct clades: one clade correlates with pyrimidine formation in the biosynthesis of bleomycin,[9] tallysomycin,[39] and zorbamycin;[40] another clade correlates with conjugate addition instead of amide bond formation in the same biosynthetic pathways; the third clade correlates with pyrrolizidine formation in the biosynthesis of azabicyclene,[10] legonmycin,[41] and brabantamide[42] (Figure S22). The remaining 17 C$_{modAA}$ domains do not form a distinctive group, most of which likely incorporate dehydroamino acids in the natural products via both dehydration of Ser/Thr and amide bond formation. Some of these C$_{modAA}$-installed dehydroamino acids likely undergo further modifications catalyzed by other biosynthetic enzymes.[38,43,44] Overall, the grouping of C$_{modAA}$ domains does not appear to follow the phylogeny of the producing bacteria, because C$_{modAA}$ domains in the same gene cluster are located in separate clades (e.g., BlmVI and BlmX, Figure S22).

We prospected for novel nonribosomal peptides that may contain dehydroamino acids. We generated a sequence similarity network (SSN)[45] of 4391 putative C$_{modAA}$ domains, which reveals large groups of C$_{modAA}$ domains that are not associated with any known natural products (Figure S23). One notable group harbors several tandem C$_{modAA}$-containing modules in a single NRPS (Figures 6C and S23). The repeat of C$_{modAA}$-containing modules suggests that the products of these NRPSs contain adjacent dehydroamino acids, which are less common in nonribosomal peptides in comparison with RiPPs (Figure S24). Interestingly, although albopeptide contains two adjacent Dha and Dhb, the C$_{modAA}$ domains in its NRPS, AlbB, are located in a different group from the main group that contains tandem C$_{modAA}$ modules in the SSN. Similarly, the C$_{modAA}$ domains in BlmVI and BlmX in bleomycin biosynthesis are located in separate groups in the SSN. These observations corroborate with our phylogenetic analysis that the clustering of C$_{modAA}$ domains does not necessarily follow the phylogeny of the producing bacterial strains. The large number of uncharacterized C$_{modAA}$-containing gene clusters suggests that dehydroamino acids may be much more prevalent in nonribosomal peptides than currently known, which presents a wide and unexplored natural product space that may be accessed using C$_{modAA}$ as a guide.

DISCUSSION

We report the biochemical and structural characterization of AmbE-C$_{modAA}$, a member of a major subfamily of C domains with a proposed function in modifying amino acids. We demonstrate that AmbE-C$_{modAA}$ catalyzes dehydrative decarboxylation of pre-Ala-AMB to Ala-AMB. We also provide bioinformatics analysis of 27 different C$_{modAA}$ domains from 20 natural product pathways, which supports a unifying function for C$_{modAA}$ domains in dehydration. Our analysis suggests that C$_{modAA}$ domains either directly incorporate dehydroamino acids into natural products or generate dehydroamino acid intermediates that enable different downstream biosynthetic transformations.

Using AmbE-C$_{modAA}$ as a model to probe C$_{modAA}$ function, we determine that AmbE-C$_{modAA}$ modifies pre-Ala-AMB on the donor T$_1$ domain and transfers the product Ala-AMB to the acceptor T$_2$ domain. Although this conclusion appears

https://doi.org/10.1021/jacs.1c13404
J. Am. Chem. Soc. 2022, 144, 14057−14070

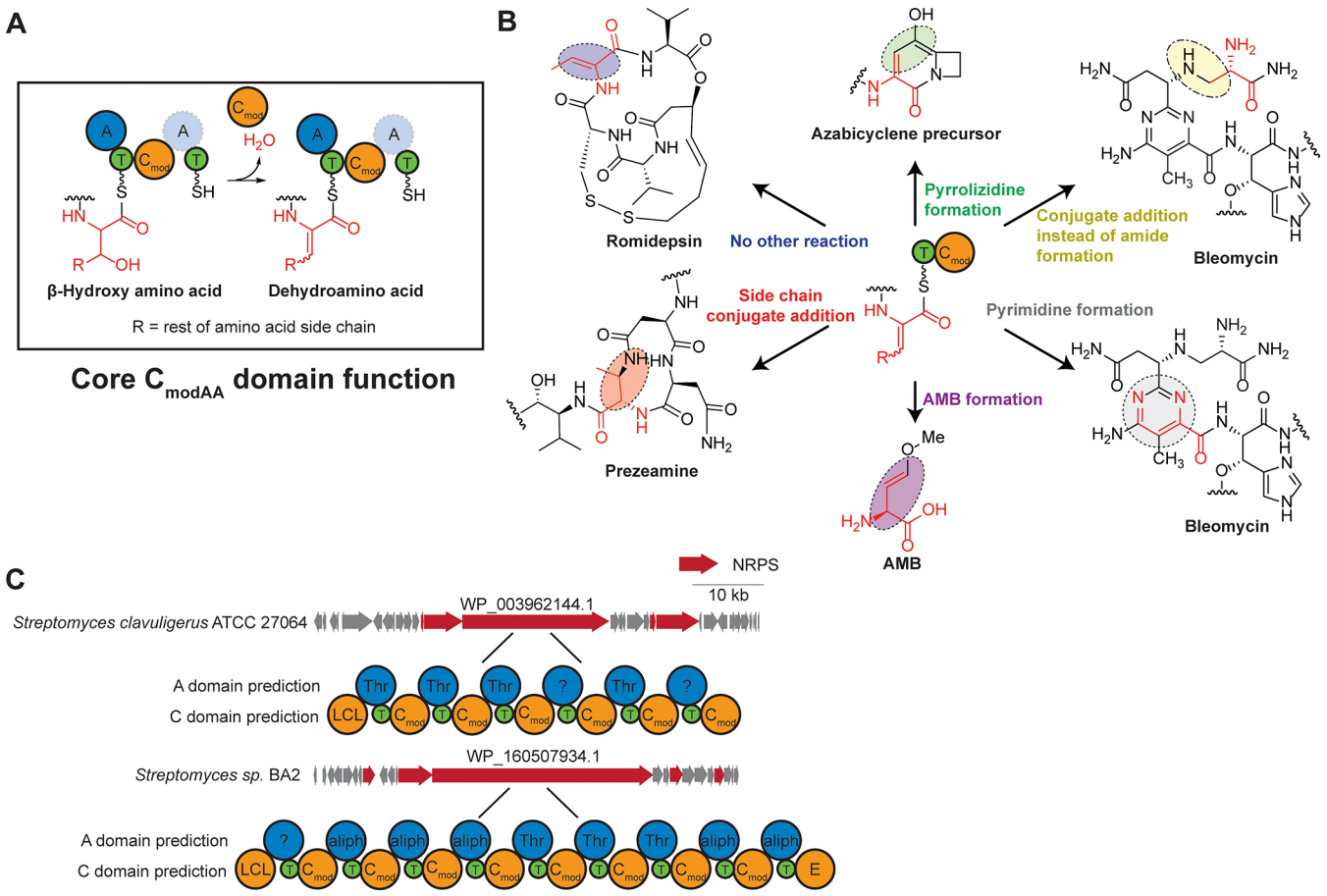


Figure 6. Bioinformatics analysis of C_{modAA} domains. (A) C_{modAA} domains are proposed to catalyze core dehydration. (B) Dehydration catalyzed by C_{modAA} domains could enable diverse modifications in nonribosomal peptide biosynthesis. (C) Uncharacterized biosynthetic gene clusters that contain tandem NRPS modules that harbor C_{modAA} domains. Aliph, aliphatic amino acid.

contradictory to our previous report that *holo* T_2 is required for the production of Ala-AMB,[16] reanalyses of the prior data revealed that the T_2 mutant of AmbE generates free Ala-AMB in the presence of cysteamine (Figure S6). This unexpected finding supports the conclusion that C_{modAA} converts pre-Ala-AMB to Ala-AMB on T_1. Minimal production of Ala-AMB by the T_2 mutant without cysteamine also indicates that transfer of Ala-AMB from T_1 to T_2 is required for release of Ala-AMB by the TE domain, unless cysteamine is added to offload Ala-AMB from AmbE into solution.

Compared to other subfamilies of C domains that also modify amino acids, including the dual condensation/epimerization domains (dual C/E) and heterocyclization domains (Cy), the timing of C_{modAA} action is reminiscent of dual C/E that catalyzes epimerization on the donor T domain prior to amide bond formation[46] but distinct from the Cy domain that catalyzes cyclization on the acceptor T domain after amide bond formation.[47,48] The recent report of two C_{modAA} domains from albopeptide biosynthesis also corroborates our finding that C_{modAA} catalyzes dehydration on the donor T domain.[17] Additionally, this work used cysteamine to capture both a Val-Ser dipeptide and a conjugate addition product between cysteamine and a Val-Dha dipeptide as biosynthetic intermediates of albopeptide (Val-Dha-Dhb). No conjugate addition products between cysteamine and Dhb were observed in this work,[17] however. Similarly, we could not capture the dehydrated pre-Ala-AMB intermediate before decarboxylation to Ala-AMB despite extensive efforts using AmbE wild-type and mutants possibly because of the steric hindrance of the bulky γ-methoxy glutamate side chain. It is also possible that AmbE-C_{modAA} catalyzes a dehydrative decarboxylation without releasing the dehydroamino acid intermediate, which remains consistent with the dehydrating function of C_{modAA} domains. Nonetheless, our work and the work on albopeptide use two different systems to provide experimental evidence to support the role of C_{modAA} domains in dehydration and answer the longstanding question of the timing of dehydration.

We obtained the crystal structure of a C_{modAA} domain, which reveals a similar overall architecture typical of the C domain family that contains an active site tunnel (Figure S9B).[26] DCA predicts a canonical binding mode of the donor T_1 with C_{modAA}, but the short 15-residue linker between C_{modAA} and T_2 prevents the binding of T_2 in the canonical acceptor T-binding position.[26] A linker of this length is not unusually short for interdomain linkers in NRPSs, but the lack of an A domain in module 2 of AmbE greatly increases the distance between the acceptor site of C_{modAA} and T_2 that this linker would be expected to span. Although it is possible that T_2 binds at the end of an auxiliary tunnel in AmbE-C_{modAA} (Figure S10B), it is more likely that T_2 accesses the canonical tunnel via dimerization (Figure 3D). We show that AmbE dimerizes upon conversion to the *holo* form. Dimerization would allow the ppant arm of T_2 in one protomer to enter the canonical acceptor tunnel of the C_{modAA} in the other protomer, which would enable transthioesterification of Ala-AMB from T_1 to T_2

https://doi.org/10.1021/jacs.1c13404
J. Am. Chem. Soc. 2022, 144, 14057−14070

(Figure 3D). Transferring Ala-AMB to T_2 would facilitate hydrolysis catalyzed by the adjacent TE domain, leading to Ala-AMB release (Figure 3D).

Very few NRPS-exclusive systems have been reported to be dimers, including the six-domain vibrobactin synthetase, VibF, and the four-domain saframycin biosynthesis protein, SfmC.[49,50] Two recent dimeric structures of NRPSs reveal dimerization architectures, including the head-to-tail homodimer of the tri-domain, FmoA3, and the depsipeptide synthetase module with an embedded ketoreductase, StsA-A-KR-T, which dimerizes through a pseudo-A_{sub} domain.[51,52] Our observation of a switch in the oligomeric state upon posttranslational phosphopantetheinylation is the first of such a report. Phosphopantetheinylation has been shown to promote other interactions, such as between the P450 enzyme NikQ and the NRPS NikP1 in nikkomycin biosynthesis;[53] therefore, self-interaction dependent on the pantetheine arm is not completely without precedence. The requirement of phosphopantetheinylation of T_2 for dimerization provides an initial clue for the overall architecture of full-length AmbE.

AmbE is not the only NRPS to have a C-T linker of ∼15 residues in length. A plot of linker length between C domains immediately followed by T domains (i.e., where the module does not contain an A domain between C and T domains) shows that while the most common linker length is ∼50 residues, almost 100 C-T didomains contain fewer than 20 residues in the linker (Figure S25). The longer linkers would be sufficient for T domain binding to the canonical acceptor site (Figure 3B), whereas as our work suggests, the proteins with shorter linkers may form dimers to allow condensation or transfer to occur. The C-T partial modules can serve unique functions in different biosynthetic pathways, although it is much less common than the canonical C-A-T. For example, the C-T in SyrE is loaded in *trans* by a distal A and T domain and enables peptide extension in syringomycin biosynthesis, whereas both domains of C-T in GliP are required for the cyclization of the diketopiperazine core in gliotoxin biosynthesis.[54−56] Since AmbE-C_{modAA} does not catalyze condensation with an acceptor amino acid, T_2 may not be involved in substrate loading or chain extension. Instead, the transfer of Ala-AMB to T_2 by C_{modAA} may bring Ala-AMB to the proximity of the TE domain for hydrolysis.

Based on sequence conservation and residue positioning in the AmbE-C_{modAA} crystal structure, we conducted mutational studies of C_{modAA} in the full-length AmbE under multiple turnover conditions and in the Q-A-MT-T_1/C_{modAA}-T_2 split system. Individual point mutations of full-length AmbE do not abolish Ala-AMB production, though the double mutation of H1497A and D1501A abolishes Ala-AMB synthesis, likely due to the collapse of the local fold. Overall, results from the split system show a similar trend to those from full-length AmbE, but mutations in the split system exhibit much more profound impacts on activity (Table S6). This phenomenon was also reported for EntF, where mutations introduced in the excised C domain of EntF more significantly impact reaction rates than the same mutations introduced in the full-length EntF.[57] The difference between the full-length AmbE and the Q-A-MT-T_1/C_{modAA}-T_2 split system may be due to the isolation of the C_{modAA} reaction from the overall AMB biosynthesis that involves at least 10 transformations. Alternatively, separation of Q-A-MT-T_1 from C_{modAA}-T_2 could weaken protein−protein interactions or destabilize the C_{modAA} domain, enhancing the impact of C_{modAA}-T_2 mutations on the reaction rate. Although the activity of the split system is less robust than that of full-length AmbE, the split system remains functional and has allowed us to measure the rate of catalysis by C_{modAA}-T_2 under kinetic conditions, which is not possible using full-length AmbE. Although T_2 is dispensable for the conversion of pre-Ala-AMB to Ala-AMB by Q-A-MT-T_1-C_{modAA}, the split system requires C_{modAA} to be linked to *holo* T_2 for activity. It is possible that *holo* T_2 helps stabilize C_{modAA} in the right conformation when C_{modAA} is separated from the rest of AmbE.

We identified four residues that are mostly conserved in C_{modAA} domains and important for activity. These residues, N1700, T1650, R1605, and H1632, are positioned in or near the active site of AmbE-C_{modAA} (Figure 4A). The N1700A mutation both increases the production of Ala-AMB in the full-length AmbE assay and increases the rate of Ala-AMB formation in the split system, and T1650A increases the rate of Ala-AMB formation in the split system (Table S6). The increased activity of these mutants could be due to the widening of the C_{modAA} active site through alanine mutation, which suggests a gatekeeping function of N1700 and T1650 (Table S7). The single mutations of conserved C_{modAA} residues that significantly augment activity could inform engineering efforts of other C_{modAA} domains. R1605A mutation decreases the rate of Ala-AMB formation in the split assay (Table S6). The model of T_1-C_{modAA} positions R1605 in contact with the phosphoester of the T_1 ppant (Figure 4A), hinting at a role in positioning the T_1 ppant for a proper substrate orientation in the active site (Table S7). H1632 is also located in the modeled donor site of the active site tunnel. Reduced activity of the H1632A mutant of full-length AmbE suggests that H1632 may also play a role in substrate positioning (Table S7). D1726 is located on the modeled acceptor T_2 domain-side loop of C_{modAA} (Figure 4A) and may help maintain the local fold for interaction with T_2 (Table S7). The preponderance of important residues on the side of C_{modAA} where it likely interacts with the donor T domain is consistent with the location of C_{modAA}-catalyzed reaction on the donor T domain.

In the HHXXXDG motif, mutation of the first His (H1496) or Asp (D1501) results in insoluble proteins in the C_{modAA}-T_2 construct, which is consistent with the structural roles that these residues are thought to play in C domain folding.[22,58] Mutation of the second His (H1497) did not significantly affect the activity of full-length AmbE but lowered the reaction rate catalyzed by C_{modAA}-T_2 to below the detection limit in the split assay (Table S6). This residue is widely accepted as important for positioning the substrate or catalyzing condensation in C domains.[18,22,59] Even so, mutation of the second His in C domains in different NRPS pathways exerts different impacts on activity, ranging from completely abolishing activity to causing modest reduction,[34,35,57,58] which suggests that its role may be compensated by other C domain residues. Similarly, minimal reduction of activity of H1497A in full-length AmbE suggests that the role of H1497 is compensated by other residues in AmbE-C_{modAA}. D1734 is positioned in the active site and opposite H1497 (Figure 4A). Like the C_{modAA}-T_2-H1497A mutant, the C_{modAA}-T_2-D1734A mutant exhibits no activity in the split system (Table S6). The abolishment of activity suggests that both H1497 and D1734 are essential for activity in the split system. The role of D1734 in substrate positioning remains to be determined. The residues E1736 and Q1747 are not conserved in C_{modAA} domains. The comparable rate of the E1736A mutant to that of WT suggests that E1736 is dispensable for C_{modAA}-T_2

https://doi.org/10.1021/jacs.1c13404
J. Am. Chem. Soc. 2022, 144, 14057−14070

activity. Q1747 is positioned near the opening of the auxiliary tunnel to the active site. While Q1747A mutation lowers the rate of Ala-AMB formation, the role of Q1747 in substrate positioning or local structure is unclear.

Because AmbE-C_{modAA} does not catalyze peptide bond formation, the role of H1497 may involve positioning the pre-Ala-AMB substrate in the AmbE-C_{modAA} active site tunnel or acting as a base that removes the α-proton (Table S7). A similar role has been proposed for the second His in the HHXXXDG motif of the NocB-C_5 domain in nocardicin biosynthesis, a $^{D}C_L$ domain distinct from C_{modAA} domains.[18,19] This residue in NocB-C_5 is not required for peptide formation but is essential for the dehydration of Ser to Dha and subsequent formation of the β-lactam ring.[18,19] Since very few other residues are broadly important for C domain function besides the second His, it is possible that not all dehydrating C domains (NocB-C_5 or C_{modAA} domains) use the same mechanism or the same residue as the catalytic base for an E1cb elimination mechanism. Any functional redundancy would also prevent the assignment of a precise residue as the catalytic base. Apart from dehydration, AmbE-C_{modAA} may catalyze subsequent decarboxylation and isomerization, the latter of which likely requires an active site residue to reprotonate the α-carbon and restore the L-stereochemistry of AMB. The identity of this residue and those involved in transthioesterification of pre-AMB from T_1 to T_2 remains to be uncovered.

Besides C_{modAA} domains and NocB-C_5, dual C/E domains preceded by A domains that activate β-hydroxy amino acids are also correlated with incorporation of dehydroamino acids,[46] although such activity has not been shown biochemically. Most of these natural products are cyclic lipopeptides isolated from pseudomonads, such as syringomycin,[60] syringopeptin,[61] nunapeptin,[62] and jessenipeptin.[63] Interestingly, the dual C/E domains that correlate with incorporation of Dha/Dhb in these pathways cannot be phylogenetically distinguished from regular dual C/E domains from the same gene clusters (Figure S26A,B). Furthermore, some dual C/E domains preceded by Ser/Thr-activating A domains do not result in the incorporation of Dha/Dhb (Figure S26B). The mechanism of dehydration by dual C/E domains remains to be explored. Compared with dual C/E and $^{D}C_L$ domains that do not typically catalyze dehydration, C_{modAA} domains appear to serve a central dehydrating function that is supported by our biochemical and bioinformatic observations.

In summary, we report the structural and functional interrogation of a dehydrating C domain (C_{modAA}) in AMB biosynthesis and begin to shed light on the molecular basis of dehydration in nonribosomal peptides by the C_{modAA} subfamily. Our work provides further bioinformatic evidence to support that C_{modAA} domains perform dehydration as a unifying step in the biosynthesis of therapeutically and ecologically important natural products. Given the diverse chemical and biosynthetic transformations that dehydroamino acids participate in, our discoveries also pave the way for using C_{modAA} domains in natural product diversification, NRPS engineering, and identification of new nonribosomal peptides.

MATERIALS AND METHODS

Purification and Crystallography of AmbE-C_{modAA}. Bacterial cultures harboring pLIC-His-AmbE-C_{modAA} were started from a single colony and grown in an LB medium that was supplemented with 100 μg/mL ampicillin. A sample of 2 mL starter culture was transferred to a 1 L LB medium that was supplemented with 100 μg/mL ampicillin. The 1 L culture was grown at 37 °C for approximately 4 h until the cell density reached an OD_{600} of 0.5−0.6 when protein expression was induced with 1 mM IPTG. The induced cultures were grown at 16 °C for 16 h, and the bacterial cells were harvested by centrifugation at 6,000 RCF. The cell pellet was resuspended for lysis in Buffer cA (50 mM Tris-Cl, 200 mM NaCl, 10 mM imidazole, pH 8.0, 2 mM β-mercaptoethanol (β-ME), pH 7.5) and lysed by sonication. The lysed cells were centrifuged using a JA25.50 rotor at 20,000 RPM for 20 min at 4 °C to remove cell debris. AmbE-C_{modAA} was purified from the supernatant via a 5 mL HiTrap IMAC FF column charged with Ni^{2+}. Buffer cA was used as the wash buffer, and Buffer cB (50 mM Tris-Cl, 200 mM NaCl, 500 mM imidazole, 2 mM β-ME, pH 7.5) was used as the elution buffer over a linear gradient. The eluted sample was then combined, concentrated, and dialyzed overnight with Buffer cC (50 mM Tris-Cl, 50 mM NaCl, 2 mM β-ME, pH 7.5) and digested with His-tagged TEV protease (1 mg per 20 mg of AmbE-C_{modAA}) at 4 °C. The digested sample was passed through the HiTrap IMAC FF column again and AmbE-C_{modAA} with the His tag removed was collected in the flowthrough, which was concentrated and separated on a MonoQ HR 16/10 column using Buffer cD (50 mM Tris-Cl, 2 mM β-ME, pH 7.5), buffer cE (50 mM Tris-Cl, 2 mM β-ME, 50 mM NaCl, pH 7.5), and buffer cF (50 mM Tris-Cl, 2 mM β-ME, 1 M NaCl, pH 7.5). Buffer cD was used to equilibrate the column, Buffer cE was used to wash the column after applying the sample, and a linear gradient from Buffer cE to Buffer cF over 100 mL was used to separate AmbE-C_{modAA} from impurities. AmbE-C_{modAA} was eluted at ∼150 mM NaCl, concentrated, and further purified via a Superdex 200 16/60 column equilibrated with Buffer cG (50 mM Tris-Cl, 200 mM NaCl, 1 mM DTT, pH 7.5). The final purified sample was concentrated, flash-frozen at −80 °C, and used for crystallography experiments.

AmbE-C_{modAA} crystallized at a concentration of 10 mg/mL in sitting-well drops after ∼5 days at 4 °C using a precipitant solution containing 100 mM bis-Tris propane, pH 6.0, 20% poly(ethylene glycol) (PEG) 3350, and 0.2 M sodium iodide. Cryoprotection was performed by dipping the crystal in a solution containing 20% 2-methyl-2,4-pentanediol (MPD), 0.2 M NaCl, 50 mM Tris-Cl, pH 7.5, 0.2 M sodium iodide, 20% PEG 3350, and 100 mM bis-Tris propane, pH 6.0 before vitrification in liquid nitrogen. The diffraction dataset was collected at the 24-ID-E beamline of the NE-CAT facility at the Advanced Photon Source (APS) in Argonne, Illinois. The data was indexed to I 121 with iMosflm[64] and scaled with AIMLESS on CCP4.[65] The data was phased by PHASER in PHENIX[66] using a homology model of the C_{modAA} domain generated by SWISS-MODEL (separating the N- and C-lobes).[67] The model was refined with COOT[68] and PHENIX refine.[66] There are two copies of AmbE-C_{modAA} in the asymmetric unit.

One-Pot Assay of Ala-AMB Production. Samples of 7.5 μM AmbB and 7.5 μM AmbE (and AmbE mutants) were incubated with 0.5 μM Sfp (the promiscuous phosphopantetheinyl transferase), 100 μM coenzyme A, 8 mM $MgCl_2$, and 50 mM KH_2PO_4 (pH 8.0) at 25 °C for 15 min to reconstitute the *holo* forms of AmbB and AmbE. In a separate reaction vessel, 3.5 μM AmbC and 6.5 μM AmbD were incubated with 100 μM $(NH_4)_2Fe(SO_4)_2$ on ice for 10 min to reconstitute the Fe(II) centers. The AmbB and AmbE reconstitution reaction was mixed with 1 mM L-alanine, 1 mM L-glutamic acid, 1 mM α-KG, 1 mM sodium ascorbate, 1 mM SAM, and 4 mM ATP, as well as AmbC and AmbD reconstituted with Fe(II), which brought the final reaction volume to 50 μL and final concentrations to 7.5 μM AmbB, 3.5 μM AmbC, 6.5 μM AmbD, 7.5 μM AmbE WT or mutants, 0.5 μM Sfp, 100 μM coenzyme A, 8 mM $MgCl_2$, 100 μM $(NH_4)_2Fe(SO_4)_2$, and 50 mM KH_2PO_4 (pH 8.0). The assay was incubated at 25 °C for 2 h, quenched with 50 μL of acetonitrile, moved to −20 °C, and incubated for 20 min to precipitate the protein components. The assays were then centrifuged at 14,000 RCF for 5 min to remove the protein precipitates. A 50 μL sample of the quenched reaction supernatant was mixed with 25 μL of 4.5 mg/mL fluorenylmethyloxycarbonyl chloride (Fmoc-Cl) in acetonitrile and 25 μL of 200 mM sodium borate buffer (pH 10.4) for Fmoc-

https://doi.org/10.1021/jacs.1c13404
J. Am. Chem. Soc. 2022, 144, 14057−14070

derivatization of the amino acid substrates and the Ala-AMB product. The derivatization reaction was mixed thoroughly for 5 min before it was centrifuged for 5 min at 13,000 RCF.

A 10 μL sample was analyzed using the "general mass spectrometry parameters (Supporting information)." Samples were injected onto a Kinetex C18 column (Phenomenex, 50 mm length, 2.6 μm particle size, and 100 Å pore size) and separated using the following method at a flow rate of 0.6 mL/min. Solvent A consisted of 0.1% formic acid in water (Fisher Scientific), and solvent B consisted of 0.1% formic acid in acetonitrile (Fisher Scientific). The mobile phase was held at 2% B for 2 min, increased from 2 to 98% B over 10 min in a linear gradient, and held at 98% B for 2 min before returning to 2% B over 1 min. The relative yield was calculated as follows: peak areas in the extracted ion chromatogram that correspond to Fmoc-Ala-AMB (425.1707 $[M + H]^+$ and 447.1527 $[M + Na]^+$) or Fmoc-Ala (312.1230 $[M + H]^+$ and 334.1050 $[M + Na]^+$) were quantified using MassHunter (Agilent). Percent conversion was calculated by dividing the Fmoc-Ala-AMB peak areas by the sum of the Fmoc-Ala and Fmoc-Ala-AMB peak areas. Each reaction for an AmbE mutant was performed 2−5 times along with a reaction for the AmbE WT as a positive control. The mean of percent conversion by the AmbE WT from a total of 23 reactions was set to 100%. Percent conversion by each AmbE mutant was normalized to the mean of the WT, and the relative percent conversion was obtained.

Because the WT and mutant groups are of unequal size (23 vs 2−5, respectively), statistical significance between the WT and mutant activities was analyzed as follows: a Shapiro test shows that WT samples do not follow a normal distribution, but all of the mutant samples are normally distributed. All samples have equal variance as determined by the Levene test. A *log* transformation of WT and mutant samples yielded normally distributed data. The statistical significance was determined using Student's *t* test on the *log*-transformed data. We also analyzed the data by directly comparing the WT and each mutant in the same experiments and performing Student's *t* test. This alternative analysis showed that the D1726A mutant and the H1497A/D1501A double mutant are significantly different from the WT ($p < 0.05$), although H1632A, N1700A, and D1696A are not significantly different from the WT controls in the same experiments. The difference between the two statistical analyses suggests that systematic errors may exist in the experiments for the one-pot assay of full-length AmbE H1632A, N1700A, and D1696A mutants. Nonetheless, systematic errors have been taken into account in our kinetic assays of the split system because the C_{modAA}-T_2 mutant was compared to WT in the same experiments. Overall, single mutations in full-length AmbE had a modest effect on activity, whereas single mutations in C_{modAA}-T_2 had a profound impact on activity.

Kinetic Assay of Ala-AMB Formation under Single Turnover Conditions. The one-pot reaction for Ala-AMB formation was modified to measure kinetics using the split AmbE-Q-A-MT-T_1 and C_{modAA}-T_2 system. Two separate initial reactions were prepared, and concentrations of each component varied based on the concentration of stock proteins and are presented as ranges. In one reaction, samples of AmbB (10.5−11.3 μM) and AmbE-Q-A-MT-T_1 (35.1−37.7 μM) were incubated with Sfp (0.35−0.38 μM), coenzyme A (70−80 μM), $MgCl_2$ (5.6−6 mM), and KH_2PO_4 (pH 8.0) (35.1−37.7 mM) at 25 °C for 10 min. To this reaction containing Q-A-MT-T_1 were added L-alanine (1.4−1.5 mM), L-glutamic acid (1.4−1.5 mM), α-KG (1.4−1.5 mM), sodium ascorbate (1.4−1.5 mM), SAM (1.4−1.5 mM), ATP (5.6−6.0 mM), $(NH_4)_2Fe(SO_4)_2$ (140−150 μM), AmbC (4.9−5.3 μM), and AmbD (9.1−9.7 μM). This reaction mixture was incubated for 45 min at 25 °C. In a separate reaction, AmbE-C_{modAA}-T_2 WT or the mutant (74−87 μM) was incubated with Sfp (0.74−0.87 μM), coenzyme A (150−170 μM), $MgCl_2$ (11.9−13.9 mM), and KH_2PO_4 (pH 8.0) (74.4−87 mM) at 25 °C for 45 min to generate *holo* C_{modAA}-T_2. The reaction mixture containing Q-A-MT-T_1 was combined with the C_{modAA}-T_2-containing reaction to a final volume of 100 μL. The final concentration of each component in the combined reaction was the following: 7.5 μM AmbB, 6.5 μM AmbC, 2.5 μM AmbD, 25 μM Q-A-MT-T_1, 25 μM C_{modAA}-T_2, 100 μM $(NH_4)_2Fe(SO_4)_2$, 0.5 μM Sfp, 100 μM coenzyme A, 1 mM L-alanine, 1 mM L-glutamic acid, 1 mM α-KG, 1 mM sodium ascorbate, 1 mM SAM, 4 mM ATP, 8 mM $MgCl_2$, and 50 mM KH_2PO_4 potassium phosphate (pH 8.0). Components present in both initial reactions were split equally between both reactions (Sfp, coenzyme A, KH_2PO_4, pH 8.0, $MgCl_2$). An aliquot of 15 μL of the combined reaction was transferred to a microcentrifuge tube every 2.5 min up to 15 min and immediately flash-frozen using liquid nitrogen. All of the aliquots were immediately quenched using 15 μL of acetonitrile upon removal from liquid nitrogen, and cysteamine hydrochloride was added to a concentration of 50 mM. The cysteamine cleavage reactions were mixed at 1000 RPM at 25 °C for 90 min. After 90 min, samples were moved to −20 °C and incubated for 20 min to precipitate the protein components. The assays were then centrifuged at 14,000 RCF for 5 min to remove the protein precipitates.

A 25 μL sample of the quenched reaction supernatant was mixed with 180 μL of 2.25 mg/mL Fmoc-Cl in acetonitrile and 50 μL of 200 mM sodium borate buffer (pH 10.4) for Fmoc-derivatization of the cysteamine-captured intermediates. The derivatization reaction was mixed thoroughly for 5 min before it was centrifuged for 5 min at 13,000 RCF. A 10 μL sample was analyzed using "general mass spectrometry methods (Supporting information)." Samples were analyzed by LC-MS as described in the "one-pot assay of Ala-AMB formation."

Size-Exclusion Chromatography-Coupled Multiangle Light Scattering (SEC-MALS) Analysis of *apo* and *holo* AmbE. A sample of 25 μM AmbE was incubated with 0.5 μM Sfp, 100 μM coenzyme A, 1 mM $MgCl_2$, and 50 mM KH_2PO_4 (pH 8.0) at 25 °C for 1 h to reconstitute *holo* AmbE. Negative controls lacked coenzyme A, Sfp, or $MgCl_2$, or all three. Reactions were flash-frozen in liquid nitrogen and stored at −80 °C until further use. For analysis by SEC-MALS (Wyatt DAWN HELEOS II light scattering instrument interfaced to an Agilent FPLC System equipped with a Superdex 200 Increase 10/300 GL, Wyatt T-rEX refractometer, and Wyatt dynamic light scattering module), 100 μL of each reaction was analyzed by running freshly prepared buffer (50 mM Tris-Cl pH 7.5, 0.5 mM TCEP, 150 mM NaCl, 0.1% w/v sodium azide) over the Superdex column for 50 min at 0.5 mL/min.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.1c13404.

Materials, additional experimental procedures, tables and supporting figures, including mass spectrometry, crystallographic, bioinformatic, and protein purity data (PDF)

Accession Codes

AmbB, AAG05693.1 | AmbC, AAG05692.1 | AmbD, AAG05691.1 | AmbE, AAG05690.1 | BlmVI, AAG02359.1 | BlmX, AAG02355.1 | AlbB, WP_150244304.1 | NocB, AAT09805.1 | VibF, ABQ21224.1 | SfmC, ABI22133.1| FmoA3, BAP16693.1 | StsA, WP_007498213.1.

■ AUTHOR INFORMATION

Corresponding Authors

Bo Li − *Department of Chemistry, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States;* orcid.org/0000-0002-8019-8891; Email: boli@email.unc.edu

T. Martin Schmeing − *Department of Biochemistry and Centre de recherche en biologie structurale, McGill University, Montréal, Canada H3G 0B1;* orcid.org/0000-0002-8061-0436; Email: martin.schmeing@mcgill.ca

Authors

Jon B. Patteson − *Department of Chemistry, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States;* orcid.org/0000-0002-5949-7678

Camille Marie Fortinez − *Department of Biochemistry and Centre de recherche en biologie structurale, McGill University, Montréal, Canada H3G 0B1*

Andrew T. Putz − *Department of Chemistry, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States*

Juan Rodriguez-Rivas − *Sorbonne Université, CNRS, Institut de Biologie Paris Seine, Biologie Computationnelle et Quantitative, LCQB, Paris 75005, France;* orcid.org/0000-0001-6262-9371

L. Henry Bryant, III − *Department of Chemistry, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States;* orcid.org/0000-0001-6141-6737

Kamal Adhikari − *Department of Biochemistry and Centre de recherche en biologie structurale, McGill University, Montréal, Canada H3G 0B1*

Martin Weigt − *Sorbonne Université, CNRS, Institut de Biologie Paris Seine, Biologie Computationnelle et Quantitative, LCQB, Paris 75005, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.1c13404

Author Contributions

‖J.B.P and C.M.F. contributed equally to this work.

Notes

The authors declare no competing financial interest.

Structure coordinates for this study have been deposited in the PDB under the accession code: 7R9X (Supporting Table 2). All data are published in the main text and supporting information. All protein sequences used for bioinformatic analysis are available at 10.5281/zenodo.6536612

■ ACKNOWLEDGMENTS

The authors thank Gary Pielak and Emily Patteson for helpful discussion, Qiang Guo for assistance with statistical analysis, Ju-Sung Kim and Ryan Mull for assistance with collecting replicate data, Sheila Podell for manually running NaPDoS on large input files, Angelos Pistofidis for generously providing reagents, and Frank Murphy and Surajit Banerjee (Advanced Photon Source) for facilitating remote collection of diffraction sets. This work was supported by the National Science Foundation (CHE1654678) and a Packard Fellowship for Science and Engineering to BL and by a CIHR Foundation Grant (FDN-148472) and a Canada Research Chair to TMS. CMF is supported by NSERC MSc and Ph.D. awards and Jon Patteson was supported by UNC-Chapel Hill Dissertation Completion Fellowship. SEC-MALS data were collected at the UNC Macromolecular Interactions Facility that is supported by a National Institutes of Health grant (P30CA016086).

■ REFERENCES

(1) Li, B.; Yu, J. P. J.; Brunzelle, J. S.; Moll, G. N.; van der Donk, W. A.; Nair, S. K. Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis. *Science* **2006**, *311*, 1464−1467.

(2) Cheng, Y.-q.; Yang, M.; Matter, A. M. Characterization of a gene cluster responsible for the biosynthesis of anticancer agent FK228 in *Chromobacterium violaceum* No. 968. *Appl. Environ. Microbiol.* **2007**, *73*, 3460−3469.

(3) Tillett, D.; Dittmann, E.; Erhard, M.; von Döhren, H.; Börner, T.; Neilan, B. A. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide−polyketide synthetase system. *Chem. Biol.* **2000**, *7*, 753−764.

(4) Bogart, J. W.; Bowers, A. A. Dehydroamino acids: chemical multi-tools for late-stage diversification. *Org. Biomol. Chem.* **2019**, *17*, 3653−3669.

(5) MacKintosh, R. W.; Dalby, K. N.; Campbell, D. G.; Cohen, P. T. W.; Cohen, P.; MacKintosh, C. The cyanobacterial toxin microcystin binds covalently to cysteine-273 on protein phosphatase 1. *FEBS Lett.* **1995**, *371*, 236−240.

(6) Dadová, J.; Galan, S. R. G.; Davis, B. G. Synthesis of modified proteins via functionalization of dehydroalanine. *Curr. Opin. Chem. Biol.* **2018**, *46*, 71−81.

(7) Repka, L. M.; Chekan, J. R.; Nair, S. K.; van der Donk, W. A. Mechanistic understanding of lanthipeptide biosynthetic enzymes. *Chem. Rev.* **2017**, *117*, 5457−5520.

(8) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **2013**, *30*, 108−160.

(9) Du, L.; Sánchez, C.; Chen, M.; Edwards, D. J.; Shen, B. The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem. Biol.* **2000**, *7*, 623−642.

(10) Patteson, J. B.; Lescallette, A. R.; Li, B. Discovery and biosynthesis of azabicyclene, a conserved nonribosomal peptide in *Pseudomonas aeruginosa*. *Org. Lett.* **2019**, *21*, 4955−4959.

(11) Walsh, C. T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat. Prod. Rep.* **2016**, *33*, 127−135.

(12) Süssmuth, R. D.; Mainz, A. Nonribosomal peptide synthesis—principles and prospects. *Angew. Chem., Int. Ed.* **2017**, *56*, 3770−3821.

(13) Challis, G. L.; Ravel, J.; Townsend, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **2000**, *7*, 211−224.

(14) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **1999**, *6*, 493−505.

(15) Ziemert, N.; Podell, S.; Penn, K.; Badger, J. H.; Allen, E.; Jensen, P. R. The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* **2012**, *7*, No. e34064.

(16) Patteson, J. B.; Dunn, Z. D.; Li, B. In vitro biosynthesis of the nonproteinogenic amino acid methoxyvinylglycine. *Angew. Chem., Int. Ed.* **2018**, *57*, 6780−6785.

(17) Wang, S.; Fang, Q.; Lu, Z.; Gao, Y.; Trembleau, L.; Ebel, R.; Andersen, J. H.; Philips, C.; Law, S.; Deng, H. Discovery and biosynthetic investigation of a new antibacterial dehydrated non-ribosomal tripeptide. *Angew. Chem., Int. Ed.* **2021**, *60*, 3229−3237.

(18) Wheadon, M. J.; Townsend, C. A. Evolutionary and functional analysis of an NRPS condensation domain integrates β-lactam, D-amino acid, and dehydroamino acid synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2026017118.

(19) Long, D. H.; Townsend, C. A. Acyl donor stringency and dehydroaminoacyl intermediates in β-lactam formation by a non-ribosomal peptide synthetase. *ACS Chem. Biol.* **2021**, *16*, 806−812.

(20) Chahtane, H.; Nogueira Füller, T.; Allard, P.-M.; Marcourt, L.; Ferreira Queiroz, E.; Shanmugabalaji, V.; Falquet, J.; Wolfender, J.-L.; Lopez-Molina, L. The plant pathogen *Pseudomonas aeruginosa* triggers a DELLA-dependent seed germination arrest in *Arabidopsis*. *eLife* **2018**, *7*, No. e37082.

(21) Belecki, K.; Townsend, C. A. Biochemical determination of enzyme-bound metabolites: preferential accumulation of a programmed octaketide on the enediyne polyketide synthase CalE8. *J. Am. Chem. Soc.* **2013**, *135*, 14339−14348.

(22) Bloudoff, K.; Schmeing, T. M. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim. Biophys. Acta, Proteins Proteomics* **2017**, *1865*, 1587−1604.
(23) Bloudoff, K.; Rodionov, D.; Schmeing, T. M. Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. *J. Mol. Biol.* **2013**, *425*, 3137−3150.
(24) Izoré, T.; Candace, Ho. Y. T.; Kaczmarski, J. A.; Gavriilidou, A.; Chow, K. H.; Steer, D. L.; Goode, R. J. A.; Schittenhelm, R. B.; Tailhades, J.; Tosin, M.; et al. Structures of a non-ribosomal peptide synthetase condensation domain suggest the basis of substrate selectivity. *Nat. Commun.* **2021**, *12*, No. 2511.
(25) Muscat, M.; Croce, G.; Sarti, E.; Weigt, M. FilterDCA: Interpretable supervised contact prediction using inter-domain coevolution. *PLoS Comput. Biol.* **2020**, *16*, No. e1007621.
(26) Reimer, J. M.; Eivaskhani, M.; Harb, I.; Guarné, A.; Weigt, M.; Schmeing, T. M. Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **2019**, *366*, No. eaaw4388.
(27) Tanovic, A.; Samel, S. A.; Essen, L.-O.; Marahiel, M. A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **2008**, *321*, 659−663.
(28) Kreitler, D. F.; Gemmell, E. M.; Schaffer, J. E.; Wencewicz, T. A.; Gulick, A. M. The structural basis of *N*-acyl-α-amino-β-lactone formation catalyzed by a nonribosomal peptide synthetase. *Nat. Commun.* **2019**, *10*, No. 3432.
(29) Drake, E. J.; Miller, B. R.; Shi, C.; Tarrasch, J. T.; Sundlov, J. A.; Allen, C. L.; Skiniotis, G.; Aldrich, C. C.; Gulick, A. M. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **2016**, *529*, 235−238.
(30) Miller, B. R.; Drake, E. J.; Shi, C.; Aldrich, C. C.; Gulick, A. M. Structures of a nonribosomal peptide synthetase module bound to MbtH-like proteins support a highly dynamic domain architecture. *J. Biol. Chem.* **2016**, *291*, 22559−22571.
(31) Tarry, M. J.; Haque, A. S.; Bui, K. H.; Schmeing, T. M. X-ray crystallography and electron microscopy of cross- and multi-module nonribosomal peptide synthetase proteins reveal a flexible architecture. *Structure* **2017**, *25*, 783−793.
(32) Tan, K.; Zhou, M.; Jedrzejczak, R. P.; Wu, R.; Higuera, R. A.; Borek, D.; Babnigg, G.; Joachimiak, A. Structures of teixobactin-producing nonribosomal peptide synthetase condensation and adenylation domains. *Curr. Res. Struct. Biol.* **2020**, *2*, 14−24.
(33) Haslinger, K.; Peschke, M.; Brieke, C.; Maximowitsch, E.; Cryle, M. J. X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. *Nature* **2015**, *521*, 105−109.
(34) Keating, T. A.; Marshall, C. G.; Walsh, C. T.; Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* **2002**, *9*, 522−526.
(35) Bergendahl, V.; Linne, U.; Marahiel, M. A. Mutational analysis of the C-domain in nonribosomal peptide synthesis. *Eur. J. Biochem.* **2002**, *269*, 620−629.
(36) Medema, M. H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J. B.; Blin, K.; de Bruijn, I.; Chooi, Y. H.; Claesen, J.; Coates, R. C.; et al. Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* **2015**, *11*, 625−631.
(37) Johnston, C. W.; Zvanych, R.; Khyzha, N.; Magarvey, N. A. Nonribosomal assembly of natural lipocyclocarbamate lipoprotein-associated phospholipase inhibitors. *ChemBioChem* **2013**, *14*, 431−435.
(38) Masschelein, J.; Clauwers, C.; Awodi, U. R.; Stalmans, K.; Vermaelen, W.; Lescrinier, E.; Aertsen, A.; Michiels, C.; Challis, G. L.; Lavigne, R. A combination of polyunsaturated fatty acid, non-ribosomal peptide and polyketide biosynthetic machinery is used to assemble the zeamine antibiotics. *Chem. Sci.* **2015**, *6*, 923−929.
(39) Tao, M.; Wang, L.; Wendt-Pienkowski, E.; George, N. P.; Galm, U.; Zhang, G.; Coughlin, J. M.; Shen, B. The tallysomycin biosynthetic gene cluster from *Streptoalloteichus hindustanus* E465-94 ATCC 31158 unveiling new insights into the biosynthesis of the bleomycin family of antitumor antibiotics. *Mol. Biosyst.* **2007**, *3*, 60−74.
(40) Galm, U.; Wendt-Pienkowski, E.; Wang, L.; George, N. P.; Oh, T.-J.; Yi, F.; Tao, M.; Coughlin, J. M.; Shen, B. The biosynthetic gene cluster of zorbamycin, a member of the bleomycin family of antitumor antibiotics, from *Streptomyces flavoviridis* ATCC 21892. *Mol. Biosyst.* **2009**, *5*, 77−90.
(41) Huang, S.; Tabudravu, J.; Elsayed, S. S.; Travert, J.; Peace, D.; Tong, M. H.; Kyeremeh, K.; Kelly, S. M.; Trembleau, L.; Ebel, R.; et al. Discovery of a single monooxygenase that catalyzes carbamate formation and ring contraction in the biosynthesis of the legonmycins. *Angew. Chem., Int. Ed.* **2015**, *54*, 12697−12701.
(42) Schmidt, Y.; van der Voort, M.; Crüsemann, M.; Piel, J.; Josten, M.; Sahl, H.-G.; Miess, H.; Raaijmakers, J. M.; Gross, H. Biosynthetic origin of the antibiotic cyclocarbamate brabantamide A (SB-253514) in plant-associated *Pseudomonas*. *ChemBioChem* **2014**, *15*, 259−266.
(43) Mori, T.; Cahn, J. K. B.; Wilson, M. C.; Meoded, R. A.; Wiebach, V.; Martinez, A. F. C.; Helfrich, E. J. N.; Albersmeier, A.; Wibberg, D.; Dätwyler, S.; et al. Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 1718−1723.
(44) Fuchs, S. W.; Grundmann, F.; Kurz, M.; Kaiser, M.; Bode, H. B. Fabclavines: bioactive peptide−polyketide-polyamino hybrids from *Xenorhabdus*. *ChemBioChem* **2014**, *15*, 512−516.
(45) Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI web resource for genomic enzymology tools: Leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* **2019**, *58*, 4169−4182.
(46) Balibar, C. J.; Vaillancourt, F. H.; Walsh, C. T. Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem. Biol.* **2005**, *12*, 1189−1200.
(47) Bloudoff, K.; Fage, C. D.; Marahiel, M. A.; Schmeing, T. M. Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 95−100.
(48) Luo, L.; Kohli, R. M.; Onishi, M.; Linne, U.; Marahiel, Ma.; Walsh, C. T. Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: Kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry* **2002**, *41*, 9184−9196.
(49) Hillson, N. J.; Balibar, C. J.; Walsh, C. T. Catalytically inactive condensation domain C1 is responsible for the dimerization of the VibF subunit of vibriobactin synthetase. *Biochemistry* **2004**, *43*, 11344−11351.
(50) Koketsu, K.; Watanabe, K.; Suda, H.; Oguri, H.; Oikawa, H. Reconstruction of the saframycin core scaffold defines dual Pictet-Spengler mechanisms. *Nat. Chem. Biol.* **2010**, *6*, 408−410.
(51) Katsuyama, Y.; Sone, K.; Harada, A.; Kawai, S.; Urano, N.; Adachi, N.; Moriya, T.; Kawasaki, M.; Shin-ya, K.; Shin-ya, K.; Senda, T. Structural and functional analyses of the tridomain-nonribosomal peptide synthetase FmoA3 for 4-methyloxazoline ring formation. *Angew. Chem., Int. Ed.* **2021**, *60*, 14554−14562.
(52) Alonzo, D. A.; Chiche-Lapierre, C.; Tarry, M. J.; Wang, J.; Schmeing, T. M. Structural basis of keto acid utilization in nonribosomal depsipeptide synthesis. *Nat. Chem. Biol.* **2020**, *16*, 493−496.
(53) Wise, C. E.; Makris, T. M. Recruitment and regulation of the non-ribosomal peptide synthetase modifying cytochrome P450 involved in nikkomycin biosynthesis. *ACS Chem. Biol.* **2017**, *12*, 1316−1326.
(54) Singh, G. M.; Vaillancourt, F. H.; Yin, J.; Walsh, C. T. Characterization of SyrC, an aminoacyltransferase shuttling threonyl and chlorothreonyl residues in the syringomycin biosynthetic assembly line. *Chem. Biol.* **2007**, *14*, 31−40.
(55) Balibar, C. J.; Walsh, C. T. GliP, a multimodular nonribosomal peptide synthetase in *Aspergillus fumigatus*, makes the diketopiperazine scaffold of gliotoxin. *Biochemistry* **2006**, *45*, 15029−15038.

https://doi.org/10.1021/jacs.1c13404
J. Am. Chem. Soc. 2022, 144, 14057−14070

(56) Baccile, J. A.; Le, H. H.; Pfannenstiel, B. T.; Bok, J. W.; Gomez, C.; Brandenburger, E.; Hoffmeister, D.; Keller, N. P.; Schroeder, F. C. Diketopiperazine formation in fungi requires dedicated cyclization and thiolation domains. *Angew. Chem., Int. Ed.* **2019**, *58*, 14589−14593.
(57) Roche, E. D.; Walsh, C. T. Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis. *Biochemistry* **2003**, *42*, 1334−1344.
(58) Stachelhaus, T.; Walsh, C. T. Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. *Biochemistry* **2000**, *39*, 5775−5787.
(59) Bloudoff, K.; Alonzo; Diego, A.; Schmeing, T. M. Chemical probes allow structural insight into the condensation reaction of nonribosomal peptide synthetases. *Cell Chem. Biol.* **2016**, *23*, 331−339.
(60) Fukuchi, N.; Isogai, A.; Nakayama, J.; Takayama, S.; Yamashita, S.; Suyama, K.; Takemoto, J. Y.; Suzuki, A. Structure and stereochemistry of three phytotoxins, syringomycin, syringotoxin and syringostatin, produced by *Pseudomonas syringae* pv. *syringae*. *J. Chem. Soc., Perkin Trans. 1* **1992**, 1149−1157.
(61) Ballio, A.; Barra, D.; Bossa, F.; Collina, A.; Grgurina, I.; Marino, G.; Moneti, G.; Paci, M.; Pucci, P.; Segre, A.; Simmaco, M. Syringopeptins, new phytotoxic lipodepsipeptides of *Pseudomonas syringae* pv. *syringae*. *FEBS Lett.* **1991**, *291*, 109−112.
(62) Michelsen Charlotte, F.; Watrous, J.; Glaring Mikkel, A.; Kersten, R.; Koyama, N.; Dorrestein Pieter, C.; Stougaard, P.; Greenberg, E. P. Nonribosomal peptides, key biocontrol components for *Pseudomonas fluorescens* In5, isolated from a greenlandic suppressive soil. *mBio 6*, e00079-15.
(63) Arp, J.; Götze, S.; Mukherji, R.; Mattern, D. J.; García-Altares, M.; Klapper, M.; Brock, D. A.; Brakhage, A. A.; Strassmann, J. E.; Queller, D. C.; et al. Synergistic activity of cosecreted natural products from amoebae-associated bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 3758.
(64) Leslie, A. G. W.; Powell, H. R. Evolving Methods for Macromolecular Crystallography. In *Processing Diffraction Data with MOSFLM*, Read, R. J.; Sussman, J. L., Eds.; Springer Netherlands: Dordrecht, 2007; pp 41−51.
(65) Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 1204−1214.
(66) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213−221.
(67) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296−W303.
(68) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486−501.

https://doi.org/10.1021/jacs.1c13404
J. Am. Chem. Soc. 2022, 144, 14057−14070