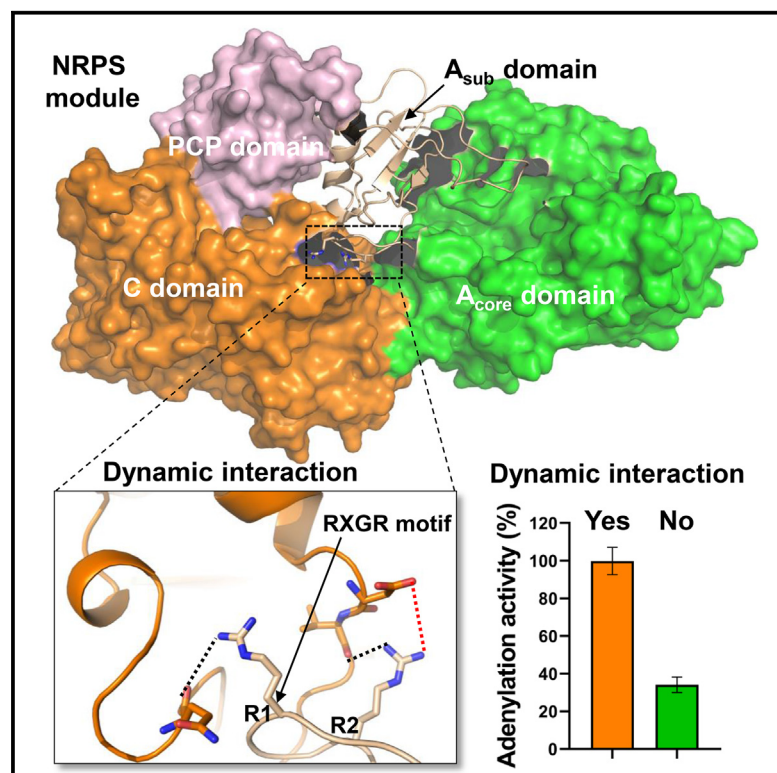# Modular catalytic activity of nonribosomal peptide synthetases depends on the dynamic interaction between adenylation and condensation domains

Graphical abstract

Authors

Ye-Jun Peng, Yuxing Chen, Cong-Zhao Zhou, Wei Miao, Yong-Liang Jiang, Xiaoli Zeng, Cheng-Cai Zhang

Correspondence

jyl@ustc.edu.cn (Y.-L.J.), zengxl@ihb.ac.cn (X.Z.), cczhang@ihb.ac.cn (C.-C.Z.)

In brief

Peng et al. analyzed the crystal structures and activity of two NRPS modules. These results provide molecular insights into the dynamic interactions between the two catalytic domains, and promote the understanding of the catalytic mechanism of NRPSs.

Highlights

- Two crystal structures representing two distinct NRPS conformations were determined

- Importance of dynamic interaction between two catalytic domains of NRPS modules

- A conserved "RGXR" motif in the adenylation domain of NRPSs mediates these interactions

⟳ CellPress

Structure

 CellPress

Article

Modular catalytic activity of nonribosomal peptide synthetases depends on the dynamic interaction between adenylation and condensation domains

Ye-Jun Peng,[1,5] Yuxing Chen,[4] Cong-Zhao Zhou,[4] Wei Miao,[2,3] Yong-Liang Jiang,[4,*] Xiaoli Zeng,[1,2,6,*] and Cheng-Cai Zhang[1,2,3,*]

[1]Key Laboratory of Algal Biology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei 430072, People's Republic of China
[2]Key Laboratory of Lake and Watershed Science for Water Security, Chinese Academy of Sciences, Nanjing 210008, People's Republic of China
[3]Hubei Hongshan Laboratory, Wuhan 430070, People's Republic of China
[4]School of Life Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230026, People's Republic of China
[5]University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China
[6]Lead contact
*Correspondence: jyl@ustc.edu.cn (Y.-L.J.), zengxl@ihb.ac.cn (X.Z.), cczhang@ihb.ac.cn (C.-C.Z.)
https://doi.org/10.1016/j.str.2024.01.010

SUMMARY

Nonribosomal peptide synthetases (NRPSs) are large multidomain enzymes for the synthesis of a variety of bioactive peptides in a modular and pipelined fashion. Here, we investigated how the condensation (C) domain and the adenylation (A) domain cooperate with each other for the efficient catalytic activity in microcystin NRPS modules. We solved two crystal structures of the microcystin NRPS modules, representing two different conformations in the NRPS catalytic cycle. Our data reveal that the dynamic interaction between the C and the A domains in these modules is mediated by the conserved "RXGR" motif, and this interaction is important for the adenylation activity. Furthermore, the "RXGR" motif-mediated dynamic interaction and its functional regulation are prevalent in different NRPSs modules possessing both the A and the C domains. This study provides new insights into the catalytic mechanism of NRPSs and their engineering strategy for synthetic peptides with different structures and properties.

INTRODUCTION

Nonribosomal peptide synthetases (NRPSs) are mega-enzymes that can assemble proteinogenic and nonproteinogenic amino acids into a vast variety of peptides, named nonribosomal peptides (NRPs).[1] Many NRPs are therapeutic agents, with antibacterial, antiviral, immunosuppressor, or antitumor activities.[2] NRPSs synthesize peptides in a modular and pipelined fashion, and each module adds one amino acid to the nascent peptide.[3,4] Therefore, the number and the specificity of the modules determine the length and the sequence of the amino acid in the corresponding peptide product.

A basic NRPS module for peptide elongation (elongation module) consists of a condensation (C) domain, an adenylation (A) domain, and a peptidyl carrier protein (PCP) domain.[1,5] The A domain is responsible for the selection and activation of the cognate amino acid, and then transfers the activated substrate to the preactivated phosphopantetheinyl (PPant) arm attached to the adjacent PCP domain.[6,7] In the next step, the PCP domain forwards the covalently bound substrate to the C domain of the same module.[6] The C domain catalyzes the peptide bond forma-

tion between the substrate from the PCP domain of the current module and the peptide chain carried on the preceding module at the active site, leading to the elongation of the peptide chain.[8] After condensation, the current PCP domain brings the elongated peptide to the C domain of the next module for another round of condensation reaction.[1,4] Usually, the initiation NRPS module in the assembly line lacks the C domain, and the termination module contains an extra thioesterase (TE) domain, which releases the peptide by cyclization or hydrolysis at the end of the whole synthesis.[9] Moreover, the diversity of NRPs may further be expanded by the action of specialized tailoring domains involved in the modification of the growing peptides, and such a domain may be part of an NRPS module or stand alone as a protein by itself.[10–12]

Biochemical and structural studies on the NRPS stand-alone domains (A, C, PCP) and NRPS didomains (A-PCP, PCP-C) indicate that the overall 3D structures of the A and C domains are conserved, but they display different conformations, suggesting that their structure changes dynamically during catalysis.[13–19] The A domain consists of a large A_{core} (~450 amino acid residues) subdomain at the N-terminus and a small A_{sub}

(~100 amino acid residues) subdomain at the C-terminus.[7,15,18] Ten conserved motifs, namely A1 to A10, were found to play important structural and catalytic roles in enzymatic reactions.[20] Three major conformations of the A domain have been reported[21]: (a) an open conformation, in which A_{sub} is rotated 30° away from the active site of A_{core}, in preparation for entry of ATP and Mg^{2+}[22]; (b) an adenylation (or closed) conformation, in which the substrate and the cofactors are bound at the active site covered by A_{sub}, allowing the adenylation reaction to occur[23]; and (c) a thiolation conformation, in which A_{sub} rotates by ~140° around A_{core}, to facilitate the attachment of the adenylated substrate to the PPant arm of the PCP domain by thiolation.[14,21] The C domain (~450 amino acids) adopts two chloramphenicol acetyltransferase folds, with the N- and C-terminal subdomains (C_N and C_C) forming a "V-shaped" pseudo dimer. The key catalytic site (HHXXXDG motif) for condensation is located at the interface of the two subdomains. For the condensation reaction, the two amino acids charged on the donor and the acceptor domains of PCP are brought together to the key catalytic site through two catalytic tunnels (donor tunnel and acceptor tunnel), located in the front and the back sides of the "V-shaped" structure, respectively.[16,24] The C domain has been proposed to exist in two major conformations: (a) an open conformation, in which no substrate is bound at the two tunnels and the "V-shaped" pseudo dimer is relatively open and (b) a closed state or condensation conformation, in which at least one of the substrates is inserted at one tunnel, and the V-shaped pseudo dimer is relatively more closed than that found in the open state.[16,24]

The increasing number of crystal structures of multidomains, modules, and cross-modules of NRPSs reported in different catalytic states highlights the importance of the dynamic interactions of the different domains during the catalytic cycle of NRPSs.[8,17,25] The PCP domain was found to interact with different catalytic domains at different regions during catalysis, and mutation of conserved residues in the observed interaction interfaces has all resulted in a loss or a reduction of the catalytic activity.[13,26–28] In addition to the dynamic domain-domain interactions related to PCP domains during catalysis, flexible interaction between adjacent A and C domains has also been found in different NRPSs. In the module or di-module crystal structures of SrfA-C,[22] AB3403,[29] ObiF,[30] and LgrA,[31] each C domain shares an extensive interaction surface with the adjacent and cognate A domain. In contrast, in the EntF termination module, the interaction surface between the A and the C domains is much reduced as compared to that found in others.[29,32] The reason behind such a difference in the interaction interfaces lies in the fact that these structures are in different catalytic states. The rotation of the PCP domain and the A_{sub} domain is necessary for delivery of the peptide intermediates to the different catalytic domains.[22,23] Recently, a few studies on NRPS modules involved in microcystin or sulfazecin production indicate that the co-existence of the adjacent A and C domains in some NRPS modules is necessary for substrate selectivity or the adenylation activity of the A domain.[33–35] These suggest the importance of the C domain in controlling dynamic domain-domain interactions in the conformational space of the flexible A_{sub}- and PCP domains.[33] However, the underlying molecular mechanism remains poorly understood.

Microcystins (MCs) are typical NRPs produced by a diverse range of bloom-forming freshwater cyanobacteria, including those belonging to the genera of *Microcystis*, *Planktothrix*, *Nostoc*, and *Oscillatoria*.[36–38] MCs are the most common hepatotoxins present in aquatic environment, which constitute a great threat to human and animal health.[39] To date, more than 270 MC variants with varying levels of toxicity have been identified,[39] and all of them share a common cyclic heptapeptide structure (D-Ala^1-X^2-D-MeAsp^3/D-Asp^3-Z^4-Adda^5-D-Glu^6-Mdha^7), in which X and Z represent variable amino acids at position 2 and 4, D-MeAsp^3 is D-*erythro*-β-methyl aspartic acid, Mdha^7 is N-methyl dehydroalanine, and Adda^5 is 3-amino-9-methoxy-2,6,8-trimethyl-10-phenyldeca-4,6-dienoic acid, respectively.[36,40] In *Microcystis aeruginosa* PCC 7806 (hereafter *Microcystis*), MCs are predicted to be synthesized by a mega-enzyme complex encoded by a 55-kb long gene cluster, the *mcy* gene cluster (Figure 1A).[41–43] The corresponding proteins McyA, McyB, and McyC are NRPSs (Figure 1B), while McyD is a PKS (polyketide synthase), and McyE and McyG are PKS/NRPS hybrids. Four proteins, McyF (amino acid racemase), McyH (ABC transporter), McyI (putative dehydrogenase), and McyJ (O-methyltransferase) are tailoring enzymes.[43] Based on the domain organization of these proteins in comparison with those of other PKSs and NPRSs, a biosynthetic pathway of MCs has been proposed.[43] *Microcystis* produces predominantly MC-LR and MC-(D-Asp^3)-LR, containing leucine (L) at position 2 and arginine (R) at position 4 (Figure 1C).[33] The synthesis of MCs starts at the Adda^5 residue. Both McyA (McyA$_1$ and McyA$_2$ modules) and McyB (McyB$_1$ and McyB$_2$ modules) consist of two NRPS modules, and are predicted to add, sequentially, amino acids Mdha^7, D-Ala^1, L-Leu^2, D-MeAsp^3/D-Asp^3 during MC synthesis (Figure 1B).[43] Since the A domain of the McyA$_2$ module is predicted to activate L-alanine (L-Ala), the epimerization domain (tailoring domain) at the C-terminus of this module was likely to be responsible for the modification of the bound L-Ala to D-Ala.[43] McyC is the terminal NRPS module responsible for adding the amino acid L-Arg onto the elongated linear peptide chain, and cyclizing and releasing the peptide by the TE domain at the C-terminal end.[43]

In this study, we found that the presence of the C domain significantly promoted the activity of the cognate A domain *in vitro* for three MC NRPS modules. Crystal structural analysis, mutagenesis of key residues, and cysteine-based crosslinking studies demonstrated that the dynamic interaction between the C and the A domains during the catalytic cycle of McyA$_2$ and McyB$_1$ modules are mediated by the "RXGR" motif and is important for the adenylation activity of the cognate A domain. Such a mechanism is prevalent in elongation and termination modules belonging to different NRPSs. These findings provide insight into the catalytic mechanism of NRPSs.

RESULTS

The C domain enhances the adenylation activity of the cognate A domain in three MC NRPS modules

A previous study suggested that the A domains of McyB$_1$ and McyC modules show the anticipated substrate activation *in vitro* only in the presence of its cognate C domain.[33] To further investigate the relationship between the C and the A domain in
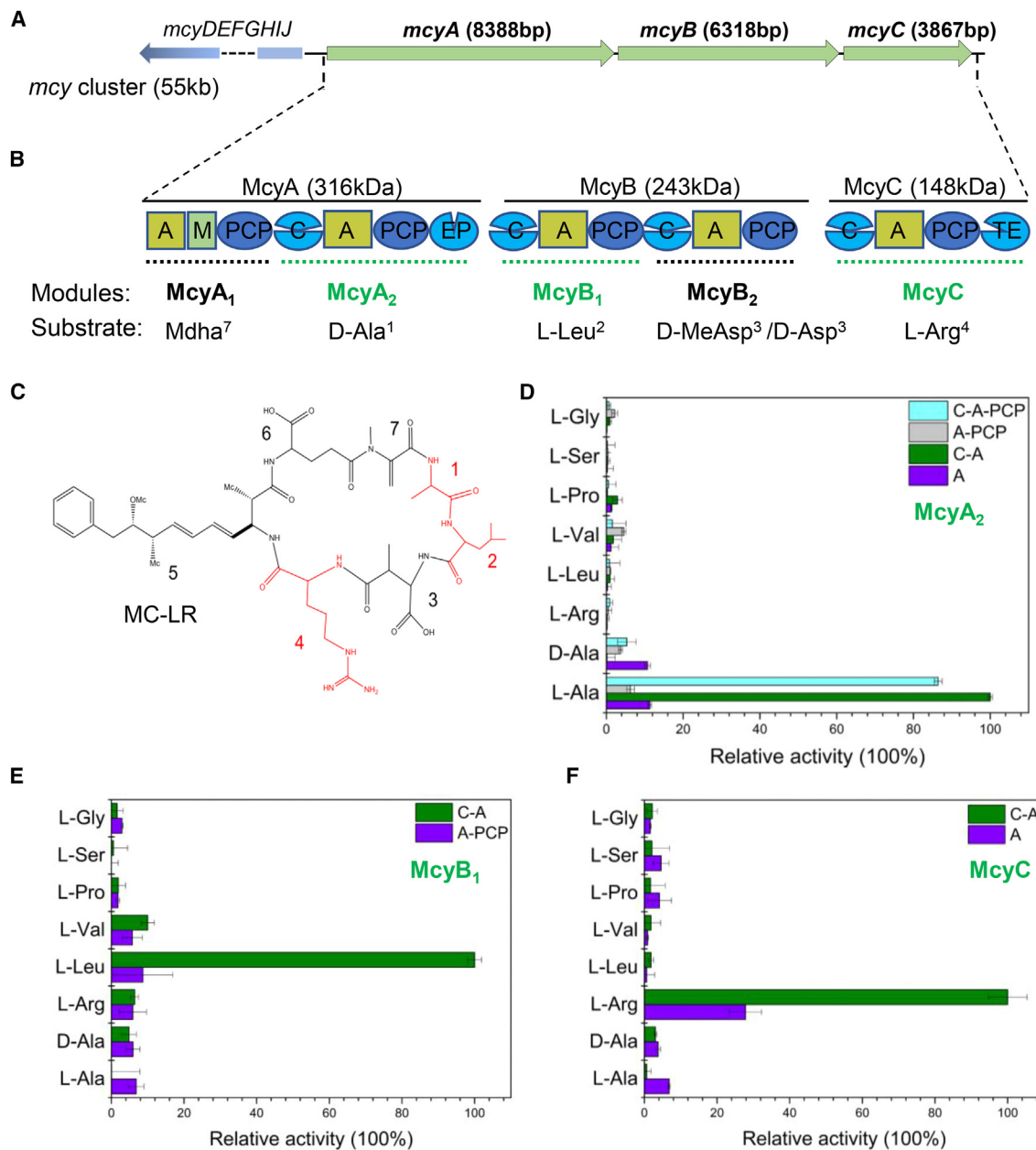
Structure
Article

CellPress

Figure 1. The presence of the C domain enhances the adenylation activity of the A domain in three MC NRPS modules

(A) Schematic representation of the MC biosynthetic gene cluster (*mcy*) in *Microcystis*.

(B) Domain organization of McyA, McyB, and McyC. A: adenylation domain, M: N-Methyltransferase domain, PCP: peptidyl carrier protein domain, C: condensation domain, EP: Epimerization domain, TE: Thioesterase domain. The NRPS modules are marker by dashed lines, and those characterized in this study by green dashed lines. The name of each module and the corresponding substrate predicted for these modules are indicated below the domain structure.

(C) Chemical structure of the cyclo-heptapeptide microcystin-LR, and the sites 1, 2, and 4 are highlighted in red.

(D) *In vitro* adenylation activities of different truncated polypeptides of McyA$_2$.

(E) *In vitro* adenylation activities of different truncated polypeptides of McyB$_1$.

(F) *In vitro* adenylation activities of different truncated polypeptides of McyC. Adenylation activities are determined using a continuous hydroxylamine release assay (see STAR Methods for details). The OD$_{650}$ value of the WT protein is set to 100%. Error bars indicate SD from the mean of two experimental replicates.

other MC NRPS modules, different modules, McyA$_1$, McyA$_2$, McyB$_1$, McyB$_2$, and McyC, were expressed in *Escherichia coli* in different combination with single A domains, C-A and A-PCP bidomains, and C-A-PCP tri-domains, respectively (STAR Methods; Figure S1A). The recombinant proteins were purified by Ni-NTA affinity chromatography followed by size exclusive chromatography. At the end, the single A domains from McyA$_2$ and McyC modules, the C-A bidomains from McyA$_2$, McyB$_1$, and McyC modules, the A-PCP bidomains from McyA$_2$ and McyB$_1$ modules, and the C-A-PCP tri-domain of McyA$_2$ module

Table 1. Crystallographic data statistics

	McyB$_1$-(C-A) (Native)[a]	McyA$_2$-(C-A-PCP) (SeMet)[a]
Data collection		
Beamline	RIGAKU MICROMAX-007	SSRF- BL19U1
Wavelength (Å)	1.5418	0.9785
Space group	$P3_12_1$	$I4_1$
Unit cell dimensions		
a, b, c (Å)	152.53, 152.53, 135.74	174.01, 174.01, 78.03
α, β, γ (°)	90.00, 90.00,120.00	90.00, 90.00, 90.00
Resolution range (Å)	13.15–2.70 (2.79–2.70)	46.55–2.50 (2.59–2.50)
R$_{meas}$	0.129 (1.549)	0.104 (0.657)
Redundancy	8.0 (6.1)	9.2 (9.7)
Completeness (%)	99.9 (98.9)	99.7(98.4)
CC$_{1/2}$	0.997 (0.679)	0.995 (0.857)
I/σ(I)	17.5 (2.2)	21.4 (3.9)
Refinement		
Resolution range (Å)	13.15–2.70	46.55–2.50
B factor (Å^2)		
Protein	66.6	43.7
Ligands	58.9	38.0
Waters	49.8	28.8
No. of reflections	47370	40,434
Completeness	98.9	99.7
R$_{work}$/R$_{free}$	0.213/0.269	0.265/0.305
No. of protein (residues)	910	1023
No. of ligand (atoms)	31	51
RMSD bond lengths (Å)	0.008	0.005
RMSD bond angles (°)	1.701	0.920
Ramachandran statistics		
Favored (%)	91.0	91.5
Allowed (%)	8.0	7.6
Outliers (%)	1.0	0.9
PDB code	8HLK	8JBR

[a]Statistics for the highest-resolution shell are shown in parentheses.

could be expressed and purified (Figure S1B). The adenylation activity of the purified proteins was tested by a continuous hydroxylamine release assay (Figure S2), using different amino acids as substrates. As shown in Figure 1 (panels D–F), the C-A bidomains of McyA$_2$, McyB$_1$, and McyC modules could activate the corresponding specific substrate with the strongest activity (L-Ala for McyA$_2$, L-Leu for McyB$_1$, L-Arg for McyC), consistent with the previous prediction.[33,43] However, the single A domains or A-PCP bidomains from the same three modules displayed a much weaker adenylation activity than the C-A bidomain proteins toward the same substrates (Figures 1D–1F). The results related to the McyB$_1$ and McyC modules were thus consistent with those reported.[33] These results indicated that

in addition to McyB$_1$ and McyC modules, the activity of the A domain in McyA$_2$ modules also depends on the presence of the cognate C domain. Furthermore, we analyzed the adenylation activity of both the A-PCP bidomain and the C-A-PCP tri-domain polypeptides from the McyA$_2$ module. We found that McyA$_2$-(C-A-PCP) also showed much stronger adenylation activity than McyA$_2$-(A-PCP) (Figure 1D). Taken together, we concluded that the presence of the C domain enhances the activity of the cognate A domain in MC NRPS modules.

Crystal structures of McyA$_2$-(C-A-PCP) and McyB$_1$-(C-A)

To gain deeper insight into the relationship between the C and A domains during catalysis in MC NRPS modules, we performed crystallization assays, and finally solved the crystal structure of McyA$_2$-(C-A-PCP) and McyB$_1$-(C-A) in the presence of the corresponding substrate L-Ala/Mg·ATP and L-Leu/Mg·ATP, with 2.45 and 2.70 Å resolutions, respectively (Table 1; STAR Methods).

The overall structure of McyB$_1$-(C-A) consists of an N-terminal C domain (residues S14-T450), and a C-terminal A domain (S473-D963) (Figure 2A). Similarly, McyA$_2$-(C-A-PCP) can be divided into an N-terminal C domain (residues Q1269-L1709), an A domain (residues C1732-N2224), and a C-terminal PCP domain (residues E2237-T2306) that is positioned near the acceptor tunnel of the C domain (Figure 2B). The overall folds of the individual C, A, and PCP domains from these two structures were consistent with previously reported structures.[4,17,25] In both structures, the C domain adopts two chloramphenicol acetyltransferase folds, with the N- and the C-terminal subdomains forming a typical "V-shaped" pseudo dimer. The active site harboring the conserved HHXXXD motif is located on a loop between the β7 strand and the α4 helix of each C domain (Figures 2A and 2B). Each A domain is subdivided into an A$_{core}$ and an A$_{sub}$ subdomains. Both A$_{core}$ subdomains are composed of two αβαβα regions that form a sandwich and a β-barrel, respectively (Figures 2A and 2B). The A$_{sub}$ subdomain in McyB$_1$-(C-A) is comprised of a central three-stranded β-sheet (β37-β39) surrounded by two helices (α26-α27) (Figure 2A), whereas only two β-sheets (β37-β38) and one helix (α27) can be well modeled in the McyA$_2$-(C-A-PCP) structure due to the weak electron density in this region (Figure 2B). The PCP domain in McyA$_2$-(C-A-PCP) consists of a four helical bundle (α29-α32), and the conserved residue S2267 for attachment of the phosphopantetheinyl (PPant) arm is located at the start of α29 (Figure 2B).

In the structure of McyB$_1$-(C-A), the electron density corresponding to L-Leu and AMP was clear in the pocket of the A domain. The sites for specific binding of L-Leu in the A domain include D668 from α20, N756 from β27, and K953 from A10 catalytic loop (Figures 2C and S3). The AMP is held in place by residues T760 from η10-α26 loop, D848 from β33, and F860 from β34 and K953. The conserved K953 interacts with the oxygen atom from the phosphate group in AMP and the carboxylate group in L-Leu, respectively (Figure 2C). In McyA$_2$-(C-A-PCP), some weak electron density could be found near the active site of the A domain, and an ATP molecule fits well here (Figure S4A). Moreover, a clear electron density was present corresponding to the PPant arm inserted into the acceptor tunnel of the C domain of McyA$_2$-(C-A-PCP) on the one hand, and covalently attached to the conserved S2267 residue from the PCP
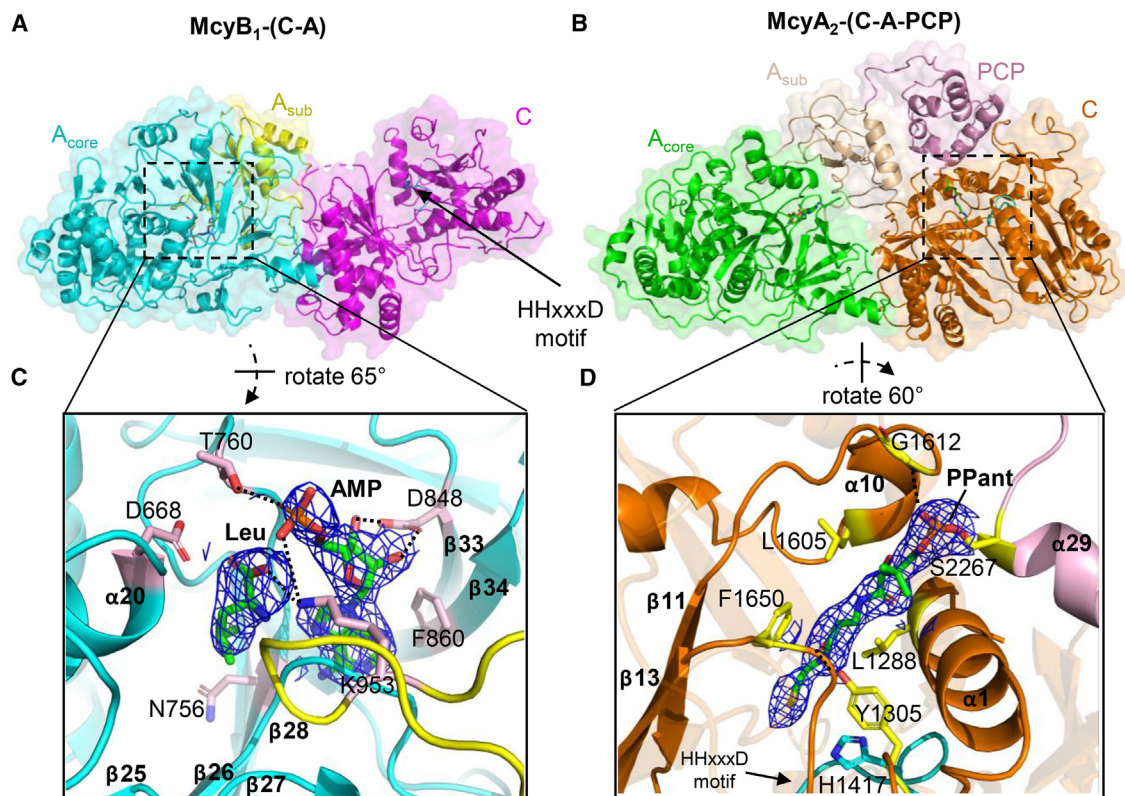
Structure
Article

 CellPress

Figure 2. Overall crystal structures of McyB$_1$-(C-A) and McyA$_2$-(C-A-PCP)

(A) Ribbon structure of McyB$_1$-(C-A), with the C domain in magenta, the A$_{core}$ domain in cyan, and the A$_{sub}$ domain in yellow. One Leu and one molecule of AMP located at the active site of the A domain are shown in green/orange-colored sticks, respectively. The "HHXXXD" motif (active site) of the C domain is indicated by black arrow.

(B) Ribbon structure of McyA$_2$-(C-A-PCP), with C domain in orange, A$_{core}$ domain in green, A$_{sub}$ domain in wheat, and PCP domain in light pink. One molecule of ATP in the A domain, a PPant arm inserted in the C domain are shown in green/orange-colored sticks. The "HXXXD" motif in the C domain is indicated by cyan.

(C) Details of binding pockets for Leu and AMP in McyB$_1$-(C-A). The Leu residue and AMP are shown in green/orange-colored sticks, superimposed with the composite omit electron density map contoured at 2.0 σ. The key polar interactions to the residues are indicated by dashed lines. The residues interacting with Leu and AMP are shown in stick representation with pink color.

(D) Interaction of PPant arm in C domain of McyA$_2$-(C-A-PCP). The PPant arm is shown in green/orange-colored sticks and are superimposed with the composite omit electron density map contoured at 1.0 σ. The PPant arm is covalently attached to Ser2267 of the PCP domain (light pink). The residues interacting with the PPant arm are shown in stick representation with yellow color and the key polar interactions are indicated by dashed line.

domain on the other (Figure 2D). The wall of the acceptor tunnel is formed by residues L1288 from α1, Tyr1305 from the linker between α1 and β2, F1650 from the loop after β13, and L1605 from α10. The interactions between the pantetheine cofactor and the side chain of these residues hold the PPant arm close to the active site (HHXXXD motif) (Figure 2D).

The structures of McyA$_2$-(C-A-PCP) and McyB$_1$-(C-A) represent two distinct NRPS conformations

The structural variations of McyA$_2$-(C-A-PCP) and McyB$_1$-(C-A) suggest that they present different conformations in the NRPS catalytic cycle. To confirm this observation, we further compared these structures with two previously well-characterized NRPSs, SrfA-C,[22] and AB3403.[29] Comparison of the four structures revealed that the C domain in McyA$_2$-(C-A-PCP) adopts a "closed" conformation similar as that of AB3403, in which the N- and C-terminal subdomains are close to each other and the PPant arm is present in the active site (Figure 3A). In contrast, the C

domain in McyB$_1$-(C-A) adopts an "open" conformation similar as that of SrfA-C, in which no substrate binding occurred at the tunnels, and the V-shaped pseudo dimers are in an open state (Figure 3A).

When the A domains of these structures were aligned, the A$_{core}$ subdomains could be well superimposed with each other, but the difference in the relative positions of the A$_{sub}$ subdomains is noticeable (Figure 3B). The A$_{sub}$ in McyB$_1$-(C-A) is closed upon the active site and well aligned with that of AB3403 (Figure 3B), indicating the A domain of McyB$_1$-(C-A) adopts the same "adenylation" conformation (or "closed" conformation) as AB3403. In contrast, the A domain of SrfA-C is in an "open" conformation, whereby its A$_{sub}$ is oriented 30° away from the active site.[22] In the case of McyA$_2$-(C-A-PCP), its A domain displays an overall "closed" conformation similar to those found in the structures of McyB$_1$-(C-A) and AB3403. However, a close-up view of this structure indicates an important rotational offset as compared to other "closed"
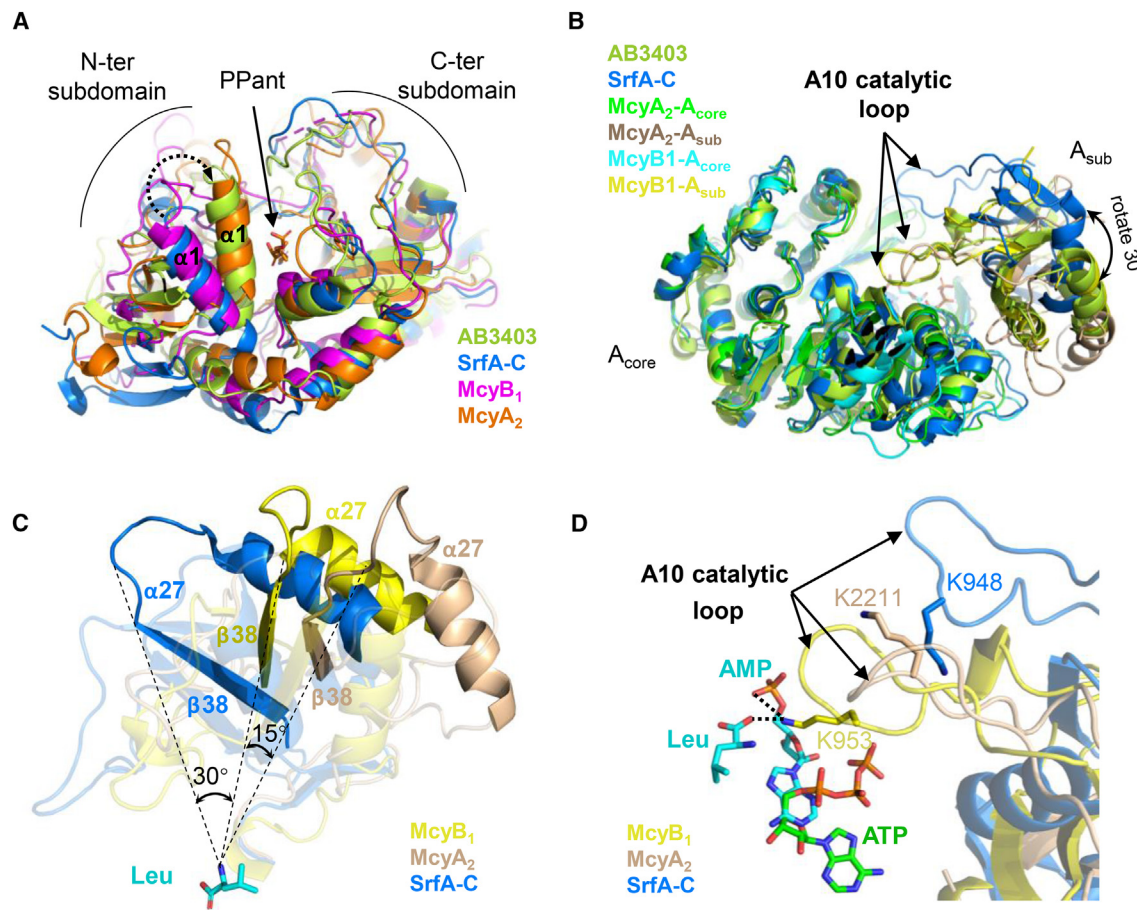
Figure 3. Structural analysis of McyB$_1$-(C-A) and McyA$_2$-(C-A-PCP)

(A) Superposition of the C domains of McyB$_1$ (magenta), McyA$_2$ (orange), SrfA-C (marine, PDB: 2VSQ), and AB3403 (limon, PDB: 4ZXH).

(B) Superposition of the A domains of McyB$_1$ (A$_{core}$, in cyan; A$_{sub}$, in yellow), McyA$_2$ (A$_{core}$, in green; A$_{sub}$, in wheat), SrfA-C (in marine), and AB3403 (in limon).

(C) Rotation angle of the A$_{sub}$ domains of McyB$_1$ (in yellow), McyA$_2$ (in wheat), and SrfA-C (in marine), relative to their respective active site (Leu as coordinate). The β38 and α27 regions of the A$_{sub}$ domains of the three structures are highlighted.

(D) Close-up view of the A10 catalytic loop of McyB$_1$ (in yellow), McyA$_2$ (in wheat), and SrfA-C (in marine) in the structural alignment. The conserved Lys residue in the A10 catalytic loop, the ATP molecule (green-orange) in McyA$_2$, and the Leu residue and the AMP molecule (cyan-orange) in McyB$_1$ are shown in stick representation. The interactions between the K953 residue and the ligands (Leu and AMP) in McyB$_1$ are indicated by dashed lines.

conformation structures (Figures 3B and S4B). These results suggest that the A domain in the structure of McyA$_2$-(C-A-PCP) could correspond to an intermediate conformation.

Indeed, by comparing the relative positions of all the A$_{sub}$ in the aligned structures, we observed a rotation by 45° relative to the catalytic sites between SrfA-C and McyA$_2$-(C-A-PCP), instead of a rotation by 30° observed between the SrfA-C ("open" conformations) and McyB$_1$-(C-A) ("adenylation" conformation). In other words, the A$_{sub}$ of McyA$_2$-(C-A-PCP) is rotated by further 15° in comparison to that of McyB$_1$-(C-A) (Figure 3C). As a consequence, the position of the critical A10 catalytic loop in the A$_{sub}$ of McyA$_2$-(C-A-PCP) is also different from those of SrfA-C and McyB$_1$-(C-A) (Figure 3D). In SrfA-C, the conserved K948 in the A10 catalytic loop moves away from the active site so that ATP and the substrate can be loaded to the active site.[22] In McyB$_1$-(C-A), the conserved K953 positions in close proximity to the carboxylate group of the L-Leu and the phosphate group of AMP, providing a catalytic impetus for the adeny-

lation reaction. However, in McyA$_2$-(C-A-PCP), the conserved residue K2211 is located in a position far away from the active site, and no substrates are present at the binding pocket, despite the fact that the A10 catalytic loop is close to the active site. Moreover, the ATP ligand is observed in a reversed conformation near, but not within the ATP-binding pocket, which is occupied by AMP in the structure of McyB$_1$-(C-A) (Figures 3D and S4A). These observations suggest that the structure of the A domain in McyA$_2$-(C-A-PCP) corresponds to an intermediate conformation between the reported "open" and "closed" conformations, in a transition from the former to the latter. Therefore, we refer to this intermediate conformation as the "pre-adenylation" conformation, which corresponds to the inception of ATP binding prior to the catalytic reaction. Compared to all published NRPS structures containing both the A and C domains,[22,29–31] the conformational combination of the A and C domains in the structures of McyA$_2$-(C-A-PCP) and McyB$_1$-(C-A) represents thus two different NRPS conformations (Figure S4C).

Structure
Article

CellPress

Dynamic interaction between the C and a domains enhances the adenylation activity of McyA$_2$-(C-A-PCP) and McyB$_1$-(C-A)

In the structures of McyA$_2$-(C-A-PCP) and that of McyB$_1$-(C-A), two different interfaces between the C and the A domains are observed, referred to as the top and the bottom interfaces, respectively (Figure 4A). For the bottom interface, the β30-β31 loop of the A$_{core}$ subdomain is stacked against the α11 and the α8-α9 loop of the C domain in both structures. Similar to other NRPSs,[22,23] this bottom interface forms a stable platform on which the PCP domain and the A$_{sub}$ domain can rearrange to promote the catalytic cycle. Although the numerous residues involved in the interaction are poorly conserved, they form super-imposable structures (Figures 4A and S3). In the top interfaces, the interactions involve the η10-α26 loop of the A$_{sub}$ domain and the α10-β11 loop together with a stretch of residues after α6 in the C domain. Although the sequences of the η10-α26 loop show a high level (67%) of similarity, the interfaces in these two structures are spatially different due to the rotation of the η10-α26 loop (Figures 4A, S3, and S5). In McyA$_2$-(C-A-PCP), two conserved residues, R2129 and R2132 of the "RGXR" motif from the η10-α26 loop, play important roles in stabilizing the top interface by forming hydrogen bonds with Q1613 and V1482, and one salt bridge with D1481, respectively (Figures 4B and S3). In contrast, in McyB$_1$-(C-A), only one salt bridge is formed between the conserved R874 and E359 at the top interface (Figure 4C).

To evaluate the functional importance of the observed interactions between the C and A$_{sub}$ domains in McyA$_2$-(C-A-PCP) and McyB$_1$-(C-A), we substituted each of the two conserved Arg residues by Ala in the "RGXR" motif and tested the effect of these mutations on the adenylation activity (STAR Methods). As shown in Figures 4D and 4E, these mutants all exhibited reductions in their adenylation activity, especially for McyA$_2$-(C-A^{R2132A}-PCP) and McyB$_1$-(C-A^{R874A}), both of which had a reduction by 60% as compared to the wild-type proteins. To rule out the possibility that such an effect was caused by a functional disruption of the A domain in McyA$_2$-(C-A^{R2132A}-PCP) and McyB$_1$-(C-A^{R874A}) mutants, we generated the variants McyA$_2$-(C^{D1481A}-A-PCP) and McyB$_1$-(C^{E359A}-A), which disrupted the same interaction, but by changing the residues on the C domains. A similar reduction of the activity using specific substrate was also observed (Figures 4D and 4E). These results indicated that the optimal adenylation activity of the A domain in McyA$_2$-(C-A-PCP) and McyB$_1$-(C-A) depends on the interactions at the top interface between the C and A$_{sub}$ domains.

We further examined the function of the interaction between the C and A domains of McyA$_2$-(C-A-PCP) and McyB$_1$-(C-A) by fixing their interface using a Cys-based cross-linking approach. To achieve an efficient interface crosslinking, a pair of residues located at the top interface in close proximity is selected and mutated into Cys. The R2129C/S1611C pair in McyA$_2$-(C-A-PCP) and the R871C/M213C pair in McyB$_1$-(C-A) were selected for the experiments (Figures 4D and 4E). When these residues were individually substituted by either alanine or cysteine, the corresponding protein variant exhibited only a marginal reduction in their adenylation activity (8%–12% reduction as compared to the wild type). However, when both the pair of residues were replaced by Cys, the corresponding proteins McyA$_2$-

(C^{S1611C}-A^{R2129C}-PCP) and McyB$_1$-(C^{M213C}-A^{R871C}) gave a 55%–65% reduction in their adenylation activity as compared to the wild type. Moreover, the reduction in this activity was reversed in both proteins by adding dithiothreitol to reduce the disulfide bonds formed by the double Cys residues (Figures 4D and 4E). These results suggest that the interaction between the C and A domains at the top interface of McyA$_2$-(C-A-PCP) and McyB$_1$-(C-A) could be fixed by disulfide bond between the double Cys substitutions, which resulted in a reduction of the adenylation activity in both proteins. However, once the fixed conformation becomes flexible again by reducing the disulfide bonds, the adenylation activity could be restored to the original levels. Taken all together, we conclude that the dynamic interaction at the top interface between the C and A$_{sub}$ domains is important for the activity of the A domain in the McyA$_2$ and McyB$_1$ modules during the catalytic cycle.

Dynamic interaction between cognate C and A domains is functionally prevalent in NRPS elongation and termination modules

Since the dynamic interaction mediated by the "RGXR" motif enhances the activity of the McyA$_2$ and McyB$_1$ modules, we wondered if similar control mechanism also exists in other NRPS modules. We first analyzed the top interface between the A and C domains in all the reported NRPS structures corresponding to distinct stages of the catalysis. The interactions observed between the η10-α26 loop of the A$_{sub}$ domain and the C domain in each respective structure mainly rely on non-bonded contacts, with notably fewer direct contacts (Figure S6). These results agree with the conclusion that the top interface between the C and A$_{sub}$ domains of an NRPS module changes dynamically during the catalytic cycle. Next, we examined the sequence conservation of the "RGXR" motif in different NRPS modules, by aligning the A8 catalytic regions surrounding the "RGXR" motif from 99 NRPS modules of different organisms (Table S3). The results indicate that the "RGXR" motif is highly conserved in the elongation and termination modules of NRPSs, but it exhibits a very low sequence similarity in the initiation modules of NRPSs or the independent A domains (Figure 5A). These results suggest that the function in adenylation activity based on dynamic interaction between the "RGXR" motif and C domain may be specifically present in the elongation and termination modules of NRPSs.

To validate this possibility, three representative NRPS modules, McyC, SrfA-C, and EntF, were selected for further investigation. SrfA-C is the termination module for surfactin synthesis in *Bacillus subtilis* ATCC 21332, and it catalyzes the addition of L-leucine to the hexapeptide precursor and releases the end product by cyclization.[22] EntF is the termination module responsible for L-serine activation during the biosynthesis of enterobactin in *E. coli*.[44] Comparison of the adenylation activity of the C-A-PCP tridomain or C-A didomain recombinant polypeptides of SrfA-C or EntF with their corresponding ones with the A domains alone revealed that the presence of the C domain in the SrfA-C and the EntF modules also enhanced their respective adenylation activity (L-Leu to SrfA-C, L-Ser to EntF) (Figure 5B). These results are thus similar to those observed with the McyA$_2$, the McyB$_1$, and the McyC modules (Figure 1). When the conserved two Arg residues in the "RGXR" motif of McyC-(C-A), SrfA-C-(C-A-PCP), and EntF-(C-A)
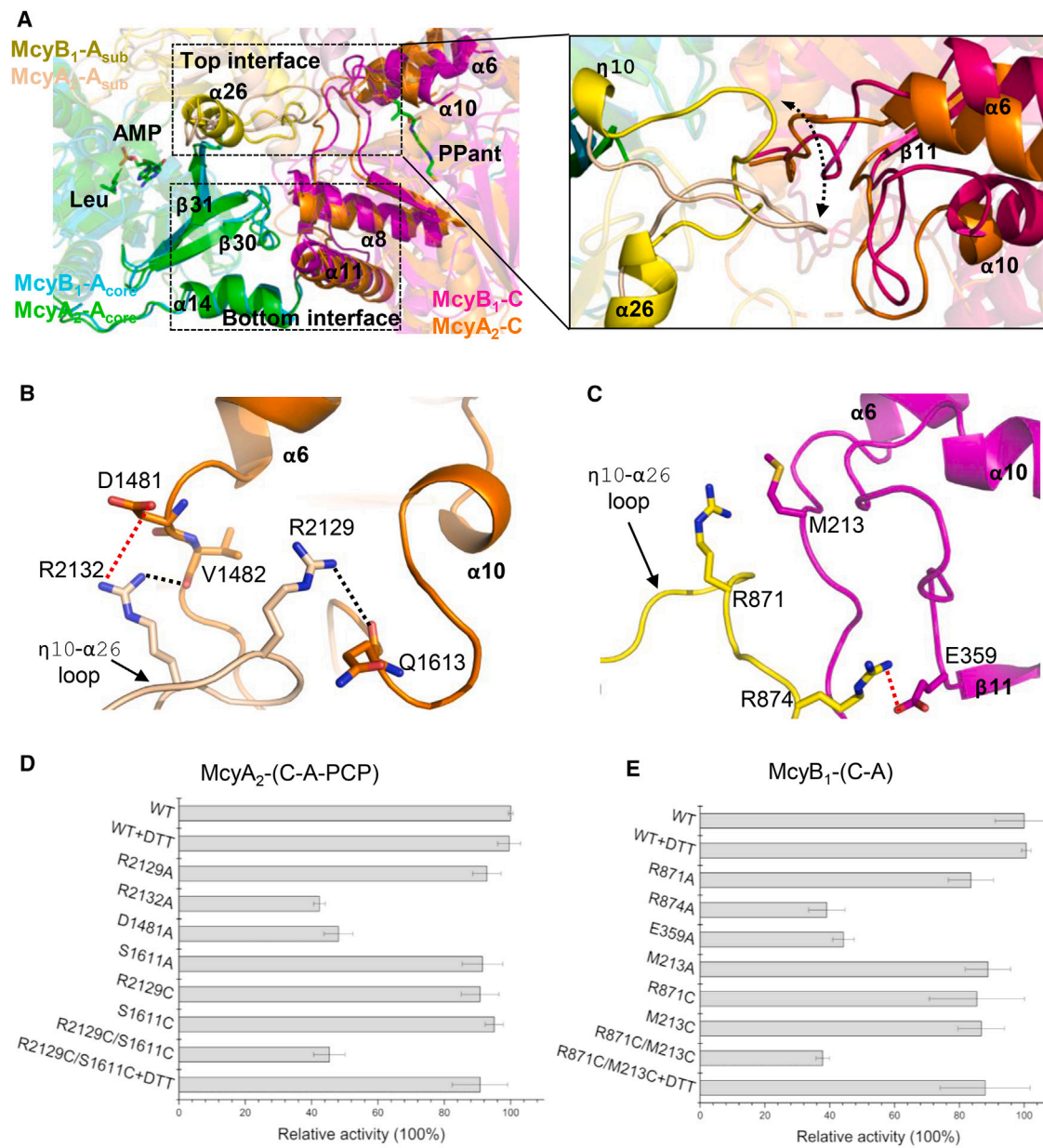
Figure 4. Structural and functional characterization of the interaction between the A and the C domains in McyB₁-(C-A) and McyA₂-(C-A-PCP)

(A) Comparison of the interaction interface between the C domain (McyB₁-C, in magenta; McyA₂-C, in orange) and the A domain (McyB₁-A$_{core}$, in cyan; McyB₁-A$_{sub}$, in yellow; McyA₂-A$_{core}$, in green; McyA₂-A$_{sub}$, in wheat) in McyB₁-(C-A) and McyA₂-(C-A-PCP). In the left panel, two interaction interfaces between the C domain and the A domain in these structures are boxed and labeled as "bottom interface" and "top interface," respectively. In the right panel, a zoom-in view of the difference at the "top interface" in McyB₁-(C-A) and McyA₂-(C-A-PCP). The rotation of the η10-α26 loop of the A$_{sub}$ domain in McyB₁-(C-A) or McyA₂-(C-A-PCP) is indicated by a dotted line with double arrows.

(B and C) Direct contact interactions between the η10-α26 loop and the C domain in the McyA₂-(C-A-PCP) structure (B) and the McyB₁-(C-A) structure (C). The salt bridge is indicated by a red dashed line, and the hydrogen bond is indicated by a black dashed line. The key residues involved in the interactions are shown in stick representation.

(D and E) The adenylation activation profile of McyA₂-(C-A-PCP) (D) and McyB₁-(C-A) (E) toward the respective specific substrate (L-Leu for McyB₁-(C-A), L-Ala for McyA₂-(C-A-PCP)) with different treatments, including single point mutation, crosslink by double Cys substitution. Adenylation activities are determined using a continuous hydroxylamine release assay. The OD$_{650}$ value of the WT protein is set at 100%. Error bars indicate standard deviation based on the mean of two experimental replicates.

Structure
Article

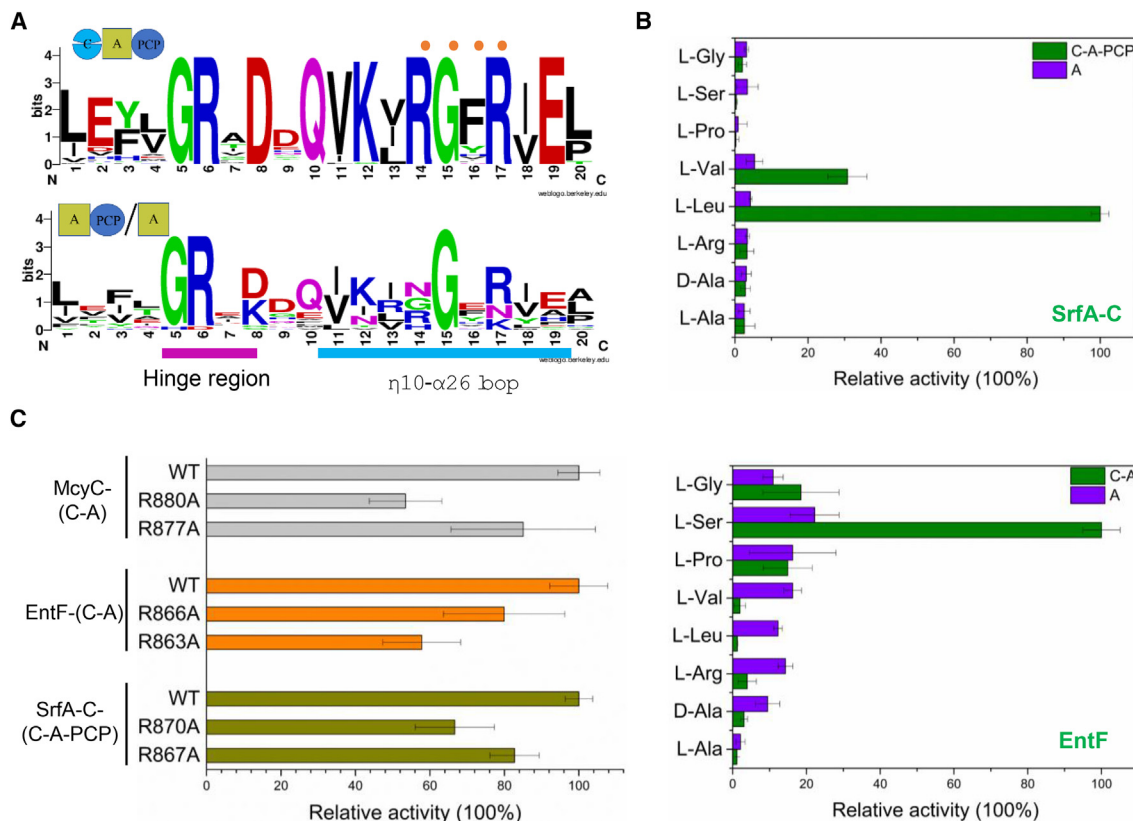CellPress

Figure 5. Functional characterization of dynamic interaction between A domain and C domains in different NRPS modules

(A) Sequence logo of the A8 catalytic motifs from different NRPS modules generated by WebLogo (http://weblogo.threeplusone.com) (Color scheme Chemistry). Heights of symbols within stack indicate the relative frequency of each amino acid residue at that position. The Logo at the upper panel was obtained through multiple alignment of the A8 catalytic motifs from 79 NRPS elongation or termination modules. The Logo at the bottom panel was obtained through multiple alignment of the A8 catalytic motifs from 21 NRPS initiation or standalone A domain modules. The conserved "hinge region" was indicated with a line in purple, and the $\eta 10\text{-}\alpha 26$ loop with a line in blue. The "RGXR" motif is indicated by orange dots.

(B) *In vitro* adenylation activities of different truncated polypeptides derived from SrfA-C (upper panel) and EntF (bottom panel).

(C) The adenylation activation profiles of McyC-(C-A), SrfA-C-(C-A-PCP), and EntF-(C-A) toward their respective specific substrates (L-Arg for McyC-(C-A), L-Leu for SrfA-C-(C-A-PCP), and L-Ser for EntF-(C-A)) with or without point mutations. The adenylation activities were determined using a continuous hydroxylamine release assay. The OD_{650} value of the WT protein is set to 100%. Error bars indicate standard deviation based on the mean of three experimental replicates.

were replaced, respectively, by Ala, the variants all exhibited a reduction in various degrees in the adenylation activity (Figure 5C). These results suggested that, similar to the McyA₂ and the McyB₁ modules, the dynamic interaction mediated by the "RGXR" motif between the C and the A_{sub} domains is also present in McyC, SrfA-C, and EntF modules and is important for the adenylation activity of these modules. Therefore, the function of the dynamic interaction between the adjacent C and A_{sub} domains mediated by the "RGXR" motif is prevalent in different NRPS elongation and termination modules.

DISCUSSION

An NRPS module contains multiple domains working together for the incorporation of a single residue into the growing peptide chain in an orderly catalytic cycle. The highly dynamic architecture of these domains is considered to be the key for efficient catalytic reactions.[17,25,45,46] Due to the dynamic nature of such multidomain proteins, it is difficult to capture the structures of

the intermediate conformations of a complete NRPS module, which limits our understanding on the structural basis of the catalytic mechanisms. In this study, we obtained two crystal structures of the NRPS involved in MC biosynthesis, one for a truncated module McyB₁-(C-A) and another for a complete elongation module McyA₂-(C-A-PCP) (Figure 2). Structural analysis and comparison with the seven reported NRPS structures that all contain both the C and the A domains show that the structures of McyA₂-(C-A-PCP) and McyB₁-(C-A) reported here represent two different conformations of NRPS in a catalytic cycle (Figures 2, 3, and S4).[22,29–31,47] For McyB₁-(C-A), the absence of the PCP domain in the module prevents the transfer of the activated substrate from the A domain to the C domain, resulting in a conformation locked at the adenylation step of the catalytic cycle (Figures 2A and 3A). A similar C-A bidomain structure of an NRPS module involved in teixobactin production was previously reported (Txo2_C₁-A₁)[47]; however, the substrate and AMP were not present in the pocket of the active site in this structure, even though the A_{sub} domain is closed upon the active site.

Thus, the reported structure of Txo2_C_1-A_1 is not a perfect one to mimic the adenylation state of the corresponding NRPS module. The presence of the substrate in the structure of McyB$_1$-(C-A) reported here may better capture the structural features of the adenylation state during the catalytic cycle. For McyA$_2$-(C-A-PCP), before setting up for crystallization, the PPant transferase (Sfp enzyme) was used to activate the PPant arm of the PCP domain.[48] Consequently, the C domain in the McyA$_2$-(C-A-PCP) structure displays a conformation in which the PPant arm of the PCP domain is docked into the cleft of the C domain, a situation similar to that for those structures of modules treated by a similar process, such as AB3403,[29] ObiF1,[30] and LgrA.[31] However, unlike the A domains of these three proteins, which are positioned in the "adenylation (closed)" conformation, the A domain in our structure is in a "pre-adenylation" conformation where the substrate and the A$_{sub}$ domain were not yet positioned for the adenylation reaction to occur (Figures 3C and 3D). This pre-adenylation structure is thus characteristic of an intermediate catalytic conformation of NRPS.

All complete module structures of NRPSs previously reported correspond to three different C–A domain conformation combinations, representing three steps in the catalytic cycle. These findings indicate that the C and A domains constitute a catalytic platform, upon which the other domains can assemble and move along.[17,22,25,29] The structures of McyA$_2$-(C-A-PCP) and McyB$_1$-(C-A) obtained in this study provide two additional distinct conformations for C–A didomains during catalysis. In all these structures, the bottom interfaces for interaction between the C domain and the A$_{core}$ domain are stable, and the dynamic changes occur mainly at the level of the top interfaces between the A$_{sub}$ and the C domains, caused by rotation of the A$_{sub}$ domain during catalysis (Figure 4A).[22,29,30] Direct contact interactions at the top interfaces are not obvious in either the McyB$_1$-(C-A) structure or other structures (Figures 4C and S6); however, we found that salt bridges and hydrogen bonds are formed between the η10-α26 loop of the A$_{sub}$ domain and the C domain in the McyA$_2$-(C-A-PCP) structure (Figure 4). This is the first time that interactions based on such strong direct contacts have been observed at the flexible top interface between the C domain and the A domain of an NRPS module. These observations suggest that certain direct interaction between the A$_{sub}$ and C domains may happen at the intermediate stages during the NRPS catalytic cycle.

Beyond expanding the interaction of C-A didomains in NRPS modules during the catalytic cycle, our studies provided a detailed functional characterization on the newly discovered interactions at this region. Previous studies found that deletion of the C domain alters the adenylation activity for the McyB$_1$ and McyC NRPS modules required for MC synthesis and the module for sulfazecin production.[33,34] Our data reported here not only provided structural insight into the participation of the C domains in the adenylation reaction, but also indicated that this effect is prevalent in elongation and termination modules of NRPSs (Figures 1 and 5). Our studies demonstrated that in these modules, dynamic interaction at the top interfaces between the C and A$_{sub}$ domains is necessary for optimal adenylation activity. Based on the direct interaction at this region observed in the structure of McyA$_2$-(C-A-PCP) and the results of sequence alignment, key residues of the "RGXR" motifs

responsible for formation of salt bridges and hydrogen bonds in the η10-α26 loop of the A$_{sub}$ domains are highly conserved in different elongation and termination modules of NRPSs (Figure 5A). Moreover, disrupting the side chain of the two conserved Arg residues in the "RGXR" motif resulted in a reduction of the adenylation activity in different NRPS modules tested (Figures 4 and 5). These results suggest that the function of the discovered interaction region is widespread in NRPSs and important for catalysis.

A significant conformational rearrangement of the A domain, in particular the A$_{sub}$ domain, is the key for an efficient catalysis during the reaction. The rotation of the A$_{sub}$ domain is accompanied by a change in the angle of the two conserved "hinge" residues (R and D) at the conserved A8 motif required for connecting the A$_{core}$ and A$_{sub}$ domains (Figure 5A).[49,50] Since the conserved η10-α26 loop is also a part of the A8 catalytic motif in the NRPS elongation and termination modules, just after the "hinge" residues (Figure 5A), it is possible that the conserved "RGXR" motif in the η10-α26 loop plays an important role in assisting the rotation of the A$_{sub}$ domain and in coordinating the catalytic reaction of the adjacent A and C domains during catalysis. Based on previous studies and our findings, we propose a model on the control mechanism of the adenylation activity in an NRPS module mediated by dynamic interaction between the two domains (Figure 6). When the substrate of an NRPS module binds at the pocket of the A$_{core}$ domain, it triggers a change at a right angle of the "hinge" residues located between the two subdomains. Subsequently, the dynamic interaction between the "RXGR" motif and the C domain guides the rotational direction of the A$_{sub}$ domain, enabling a more efficient state transition and a high adenylation efficiency of the A domain. When the dynamic interaction between the "RXGR" motif and the C domain was disrupted, the A$_{sub}$ domain rotates in a random fashion, thereby resulting in a low adenylation efficiency of the A domain in an NRPS module.

Certain details in this model still need to be understood, in particular more intermediate structural snapshots are required for a comprehensive understanding of the catalytic cycle. Nevertheless, the model provides a framework for the underlying control mechanism involved in the catalytic activity in the presence of a substrate in NRPS modules. NRPSs are attractive targets for bioengineering for peptide products in the era of synthetic biology. The elucidation of the control mechanism of the catalytic activity suggested that the bioactivities and the yield of bioengineered NRPs could be improved by swapping paired C-A domains from different NRPSs modules, enabling the production of synthetic peptides with new amino acid substrates. Such synthetic peptides may be endowed with new structures and properties.

Significance

NRPs include a variety of bioactive compounds, and are important targets for drug development. NRPs are synthetized by NRPSs, which are large multidomain enzymes and are among the most complex enzymatic systems in biology. The modular structure of NRPSs could be explored for generating new bioactive compounds by engineering or modifying the synthetases modules or domains. However, such a strategy for drug development is hardly possible because of the complexity in the catalytic
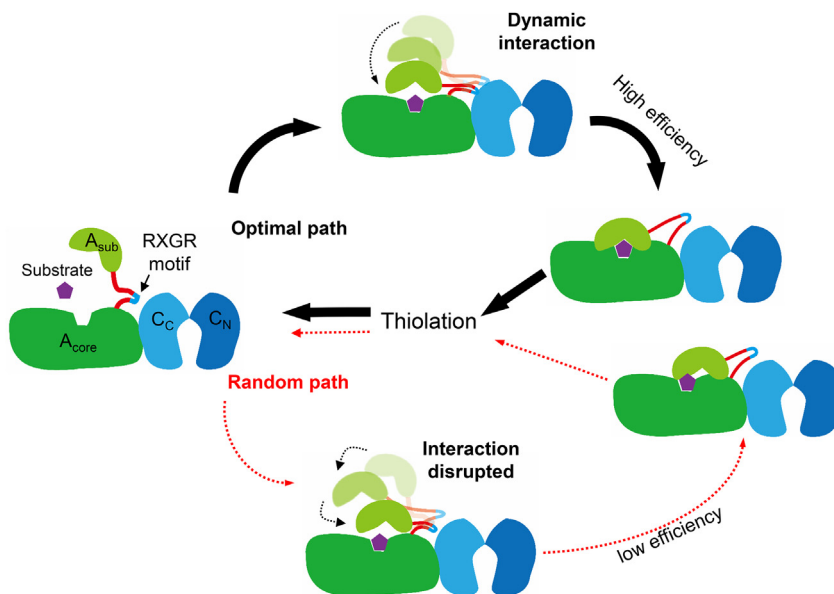
Structure
Article

    CellPress

Figure 6. A model for the control mechanism of the adenylation activity in an NRPS module, mediated by dynamic interaction between the A and C domains

Binding of the substrate in an NRPS module triggers a change of the "hinge" region (red line) located at the linker between the A_{sub} domain and the A_{core} domain. Subsequently, dynamic interaction between the "RXGR" motif which after the hinge region and the C domain guide the rotation of the A_{sub} domain in a direction that allows an efficient state transition (from the open conformation to adenylation/closed conformation) of the A domain. In this case, the NRPS module shows a highly efficient adenylation activity (optimal path, upper part, with thick arrows). When the interaction between the "RXGR" motif and the C domain was disrupted, the rotation of the A_{sub} domain occurs rather randomly with ill fitted state transition or conformational changes, and thereby, the A domain in the NRPS module shows adenylation activity with poor efficiency.

mechanism determined by each module and the interaction between two adjacent modules. In this study, by using structural biology and biochemical characterization, we identified and characterized the role of the C domain in helping the A domain to activate the substrate during catalysis of MC NRPSs. Such a mechanism is highly conserved in different NRPS modules tested, which provides not only new insight for understanding of the adenylation reaction mechanism, but also for NRPSs engineering.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Microbe strains
- METHOD DETAILS
  - Plasmids construction
  - Protein expression and purification
  - Crystallization
  - Structure determination of $McyA_2$-(C-A-PCP) and $McyB_1$-(C-A)
  - Continuous hydroxylamine release assay
  - Bioinformatic analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.str.2024.01.010.

AUTHOR CONTRIBUTIONS

Conceptualization, C.-C.Z., C.-Z.Z., Y.C., and X.Z.; methodology, Y.-J.P. and Y.-L.J.; investigation, Y.-J.P.; formal analysis, Y.-J.P., X.Z., and Y.-L.J.; visualization, Y.-J.P. and X.Z.; writing – original draft, X.Z. and Y.-J.P.; writing – review & editing, X.Z., Y.-L.J., C.-Z.Z., Y.C., W.M., and C.-C.Z.; funding acquisition, C.-C.Z. and X.Z.

REFERENCES

1. Strieker, M., Tanović, A., and Marahiel, M.A. (2010). Nonribosomal peptide synthetases: structures and dynamics. Curr. Opin. Struct. Biol. *20*, 234–240. https://doi.org/10.1016/j.sbi.2010.01.009.

2. Felnagle, E.A., Jackson, E.E., Chan, Y.A., Podevels, A.M., Berti, A.D., McMahon, M.D., and Thomas, M.G. (2008). Nonribosomal peptide synthetases involved in the production of medically relevant natural products. Mol. Pharm. *5*, 191–211. https://doi.org/10.1021/mp700137g.

3. Fischbach, M.A., and Walsh, C.T. (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem. Rev. *106*, 3468–3496. https://doi.org/10.1021/cr0503097.

4. Koglin, A., and Walsh, C.T. (2009). Structural insights into nonribosomal peptide enzymatic assembly lines. Nat. Prod. Rep. *26*, 987–1000. https://doi.org/10.1039/b904543k.

5. Weber, T., and Marahiel, M.A. (2001). Exploring the domain Structure of modular nonribosomal Peptide synthetases. Structure *9*, R3–R9. https://doi.org/10.1016/S0969-2126(00)00560-8.

6. Crosby, J., and Crump, M.P. (2012). The structural role of the carrier protein – active controller or passive carrier. Nat. Prod. Rep. *29*, 1111–1137. https://doi.org/10.1039/c2np20062g.

7. Gulick, A.M. (2009). Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. ACS Chem. Biol. *4*, 811–827. https://doi.org/10.1021/cb900156h.

8. Wang, J., Li, D., Chen, L., Cao, W., Kong, L., Zhang, W., Croll, T., Deng, Z., Liang, J., and Wang, Z. (2022). Catalytic trajectory of a dimeric nonribosomal peptide synthetase subunit with an inserted epimerase domain. Nat. Commun. *13*, 592. https://doi.org/10.1038/s41467-022-28284-x.

9. Lawson, D.M., Derewenda, U., Serre, L., Ferri, S., Szittner, R., Wei, Y., Meighen, E.A., and Derewenda, Z.S. (1994). Structure of a myristoyl-ACP-specific thioesterase from Vibrio harveyi. Biochemistry *33*, 9382–9388. https://doi.org/10.1021/bi00198a003.

10. Süssmuth, R.D., and Mainz, A. (2017). Nonribosomal peptide synthesis—principles and prospects. Angew. Chem., Int. Ed. Engl. *56*, 3770–3821. https://doi.org/10.1002/anie.201609079.

11. Walsh, C.T., Chen, H., Keating, T.A., Hubbard, B.K., Losey, H.C., Luo, L., Marshall, C.G., Miller, D.A., and Patel, H.M. (2001). Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. Curr. Opin. Chem. Biol. *5*, 525–534. https://doi.org/10.1016/S1367-5931(00)00235-0.

12. Konz, D., and Marahiel, M.A. (1999). How do peptide synthetases generate structural diversity? Chem. Biol. *6*, R39–R48. https://doi.org/10.1016/S1074-5521(99)80002-7.

13. Samel, S.A., Schoenafinger, G., Knappe, T.A., Marahiel, M.A., and Essen, L.-O. (2007). Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. Structure *15*, 781–792. https://doi.org/10.1016/j.str.2007.05.008.

14. Reger, A.S., Wu, R., Dunaway-Mariano, D., and Gulick, A.M. (2008). Structural characterization of a 140° domain movement in the two-step reaction catalyzed by 4-lhlorobenzoate: CoA ligase. Biochemistry *47*, 8016–8025. https://doi.org/10.1021/bi800696y.

15. Mitchell, C.A., Shi, C., Aldrich, C.C., and Gulick, A.M. (2012). Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. Biochemistry *51*, 3252–3263. https://doi.org/10.1021/bi300112e.

16. Bloudoff, K., Rodionov, D., and Schmeing, T.M. (2013). Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. J. Mol. Biol. *425*, 3137–3150. https://doi.org/10.1016/j.jmb.2013.06.003.

17. Patel, K.D., MacDonald, M.R., Ahmed, S.F., Singh, J., and Gulick, A.M. (2023). Structural advances toward understanding the catalytic activity and conformational dynamics of modular nonribosomal peptide synthetases. Nat. Prod. Rep. *40*, 1550–1582. https://doi.org/10.1039/d3np00003f.

18. Tan, X.F., Dai, Y.N., Zhou, K., Jiang, Y.L., Ren, Y.M., Chen, Y., and Zhou, C.Z. (2015). Structure of the adenylation-peptidyl carrier protein didomain of the Microcystis aeruginosa microcystin synthetase McyG. Acta Crystallogr. D Biol. Crystallogr. *71*, 873–881. https://doi.org/10.1107/S1399004715001716.

19. Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. EMBO J. *16*, 4174–4183. https://doi.org/10.1093/emboj/16.14.4174.

20. Turgay, K., Krause, M., and Marahiel, M.A. (1992). Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. Mol. Microbiol. *6*, 2743–2744. https://doi.org/10.1111/j.1365-2958.1992.tb01451.x.

21. Yonus, H., Neumann, P., Zimmermann, S., May, J.J., Marahiel, M.A., and Stubbs, M.T. (2008). Crystal structure of DltA implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. J. Biol. Chem. *283*, 32484–32491. https://doi.org/10.1074/jbc.M800557200.

22. Tanovic, A., Samel, S.A., Essen, L.-O., and Marahiel, M.A. (2008). Crystal structure of the termination module of a nonribosomal peptide synthetase. Science *321*, 659–663. https://doi.org/10.1126/science.1159850.

23. Reimer, J.M., Aloise, M.N., Harrison, P.M., and Schmeing, T.M. (2016). Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. Nature *529*, 239–242. https://doi.org/10.1038/nature16503.

24. Keating, T.A., Marshall, C.G., Walsh, C.T., and Keating, A.E. (2002). The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. Nat. Struct. Biol. *9*, 522–526. https://doi.org/10.1038/nsb810.

25. Izoré, T., and Cryle, M.J. (2018). The many faces and important roles of protein-protein interactions during non-ribosomal peptide synthesis. Nat. Prod. Rep. *35*, 1120–1139. https://doi.org/10.1039/c8np00038g.

26. Drake, E.J., Nicolai, D.A., and Gulick, A.M. (2006). Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain. Chem. Biol. *13*, 409–419. https://doi.org/10.1016/j.chembiol.2006.02.005.

27. Koglin, A., Mofid, M.R., Löhr, F., Schäfer, B., Rogov, V.V., Blum, M.-M., Mittag, T., Marahiel, M.A., Bernhard, F., and Dötsch, V. (2006). Conformational switches modulate protein interactions in peptide antibiotic synthetases. Science *312*, 273–276. https://doi.org/10.1126/science.1122928.

28. Frueh, D.P., Arthanari, H., Koglin, A., Vosburg, D.A., Bennett, A.E., Walsh, C.T., and Wagner, G. (2008). Dynamic thiolation–thioesterase structure of a non-ribosomal peptide synthetase. Nature *454*, 903–906. https://doi.org/10.1038/nature07162.

29. Drake, E.J., Miller, B.R., Shi, C., Tarrasch, J.T., Sundlov, J.A., Allen, C.L., Skiniotis, G., Aldrich, C.C., and Gulick, A.M. (2016). Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. Nature *529*, 235–238. https://doi.org/10.1038/nature16163.

30. Kreitler, D.F., Gemmell, E.M., Schaffer, J.E., Wencewicz, T.A., and Gulick, A.M. (2019). The structural basis of N-acyl-α-amino-β-lactone formation catalyzed by a nonribosomal peptide synthetase. Nat. Commun. *10*, 3432. https://doi.org/10.1038/s41467-019-11383-7.

31. Reimer, J.M., Eivaskhani, M., Harb, I., Guarné, A., Weigt, M., and Schmeing, T.M. (2019). Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. Science *366*, eaaw4388. https://doi.org/10.1126/science.aaw4388.

32. Miller, B.R., Drake, E.J., Shi, C., Aldrich, C.C., and Gulick, A.M. (2016). Structures of a nonribosomal peptide synthetase module bound to MbtH-like proteins support a highly dynamic domain architecture. J. Biol. Chem. *291*, 22559–22571. https://doi.org/10.1074/jbc.M116.746297.

33. Meyer, S., Kehr, J.C., Mainz, A., Dehm, D., Petras, D., Süssmuth, R.D., and Dittmann, E. (2016). Biochemical dissection of the natural diversification of Microcystin provides lessons for synthetic biology of NRPS. Cell Chem. Biol. *23*, 462–471. https://doi.org/10.1016/j.chembiol.2016.03.011.

34. Li, R., Oliver, R.A., and Townsend, C.A. (2017). Identification and characterization of the sulfazecin monobactam biosynthetic gene cluster. Cell Chem. Biol. *24*, 24–34. https://doi.org/10.1016/j.chembiol.2016.11.010.

35. Kaniusaite, M., Tailhades, J., Marschall, E.A., Goode, R.J.A., Schittenhelm, R.B., and Cryle, M.J. (2019). A proof-reading mechanism for non-proteinogenic amino acid incorporation into glycopeptide antibiotics. Chem. Sci. *10*, 9466–9482. https://doi.org/10.1039/c9sc03678d.

36. Rinehart, K.L., Namikoshi, M., and Choi, B.W. (1994). Structure and biosynthesis of toxins from blue-green-algae (Cyanobacteria). J. Appl. Phycol. *6*, 159–176. https://doi.org/10.1007/BF02186070.

37. van Apeldoorn, M.E., van Egmond, H.P., Speijers, G.J.A., and Bakker, G.J.I. (2007). Toxins of cyanobacteria. Mol. Nutr. Food Res. *51*, 7–60. https://doi.org/10.1002/mnfr.200600185.

38. Duy, T.N., Lam, P.K., Shaw, G.R., and Connell, D.W. (2000). Toxicology and risk assessment of freshwater cyanobacterial (blue-green algal) toxins in water. Rev. Environ. Contam. Toxicol. *163*, 113–185. https://doi.org/10.1007/978-1-4757-6429-1_3.

Structure
Article

CellPress

39. Massey, I.Y., Al Osman, M., and Yang, F. (2022). An overview on cyanobacterial blooms and toxins production: their occurrence and influencing factors. Toxins *41*, 326–346. https://doi.org/10.1080/15569543.2020.1843060.

40. Botes, D.P., Tuinman, A.A., Wessels, P.L., Viljoen, C.C., Kruger, H., Williams, D.H., Santikarn, S., Smith, R.J., and Hammond, S.J. (1984). The structure of cyanoginosin-LA, a cyclic heptapeptide toxin from the cyanobacterium Microcystis aeruginosa. J. Chem. Soc. Perkin 1 *1*, 2311. https://doi.org/10.1039/P19840002311.

41. Meißner, K., Dittmann, E., and Börner, T. (1996). Toxic and non-toxic strains of the cyanobacterium Microcystis aeruginosacontain sequences homologous to peptide synthetase genes. FEMS Microbiol. Lett. *135*, 295–303. https://doi.org/10.1111/j.1574-6968.1996.tb08004.x.

42. Dittmann, E., Neilan, B.A., Erhard, M., von Döhren, H., and Börner, T. (1997). Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium Microcystis aeruginosa PCC 7806. Mol. Microbiol. *26*, 779–787. https://doi.org/10.1046/j.1365-2958.1997.6131982.x.

43. Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T., and Neilan, B.A. (2000). Structural organization of microcystin biosynthesis in Microcystis aeruginosa PCC7806: an integrated peptide–polyketide synthetase system. Chem. Biol. *7*, 753–764. https://doi.org/10.1016/S1074-5521(00)00021-1.

44. Miller, B.R., Drake, E.J., Shi, C., Aldrich, C.C., and Gulick, A.M. (2016). Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture. J. Biol. Chem. *291*, 22559–22571. https://doi.org/10.1074/jbc.M116.746297.

45. Bonhomme, S., Dessen, A., and Macheboeuf, P. (2021). The inherent flexibility of type I non-ribosomal peptide synthetase multienzymes drives their catalytic activities. Open Biol. *11*, 200386. https://doi.org/10.1098/rsob.200386.

46. Winn, M., Fyans, J.K., Zhuo, Y., and Micklefield, J. (2016). Recent advances in engineering nonribosomal peptide assembly lines. Nat. Prod. Rep. *33*, 317–347. https://doi.org/10.1039/c5np00099h.

47. Tan, K., Zhou, M., Jedrzejczak, R.P., Wu, R., Higuera, R.A., Borek, D., Babnigg, G., and Joachimiak, A. (2020). Structures of Teixobactin-producing Nonribosomal Peptide Synthetase Condensation and Adenylation Domains. Curr. Res. Struct. Biol. *2*, 14–24. https://doi.org/10.1016/j.crstbi.2020.01.002.

48. Quadri, L.E., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P., and Walsh, C.T. (1998). Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. Biochemistry *37*, 1585–1595. https://doi.org/10.1021/bi9719861.

49. Wu, R., Reger, A.S., Lu, X., Gulick, A.M., and Dunaway-Mariano, D. (2009). The mechanism of domain alternation in the acyl-adenylate forming ligase superfamily member 4-lhlorobenzoate: Coenzyme A ligase. Biochemistry *48*, 4115–4125. https://doi.org/10.1021/bi9002327.

50. Reger, A.S., Carney, J.M., and Gulick, A.M. (2007). Biochemical and crystallographic analysis of substrate binding and conformational changes in Acetyl-CoA synthetase. Biochemistry *46*, 6536–6546. https://doi.org/10.1021/bi6026506.

51. Walden, H. (2010). Selenium incorporation using recombinant techniques. Acta Crystallogr. D Biol. Crystallogr. *66*, 352–357. https://doi.org/10.1107/S0907444909038207.

52. Zeng, X., Huang, M., Sun, Q.X., Peng, Y.J., Xu, X., Tang, Y.B., Zhang, J.Y., Yang, Y., and Zhang, C.C. (2023). A c-di-GMP binding effector controls cell size in a cyanobacterium. Proc. Natl. Acad. Sci. USA *120*, e2221874120. https://doi.org/10.1073/pnas.2221874120.

53. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326. https://doi.org/10.1016/S0076-6879(97)76066-X.

54. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242. https://doi.org/10.1107/S0907444910045749.

55. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

56. Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. *67*, 355–367. https://doi.org/10.1107/S0907444911001314.

57. Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132. https://doi.org/10.1107/S0907444904019158.

58. Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.-W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr. *58*, 1948–1954. https://doi.org/10.1107/S0907444902016657.

59. Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291. https://doi.org/10.1107/S0021889892009944.

60. Chen, V.B., Arendall, W.B., III, 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. https://doi.org/10.1107/S0907444909042073.

61. Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr. D Biol. Crystallogr. *67*, 271–281. https://doi.org/10.1107/S0907444910048675.

62. Duckworth, B.P., Wilson, D.J., and Aldrich, C.C. (2016). Measurement of nonribosomal peptide synthetase adenylation domain activity using a continuous hydroxylamine release assay. Methods Mol. Biol. *1401*, 53–61. https://doi.org/10.1007/978-1-4939-3375-4_3.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Phosphate colorimetric assay kit	Sigma-Aldrich	Cat#MAK030
Hydroxylamine hydrochloride	Sigma-Aldrich	Cat#55460
Adenosine 5′-triphosphate disodium salt (ATP)	Sigma-Aldrich	Cat#A2383
Inorganic pyrophosphatase	Sigma-Aldrich	Cat#I1643
4′-phosphopantetheinyl transferase (Sfp)	Novoprotein	Cat#CAA44858.1
Coenzyme A hydrate (CoA)	Sigma-Aldrich	Cat#C4282
His-tagged $McyB_1$-(C-A-PCP)	This paper	N/A
His-tagged $McyB_1$-(C-A)	This paper	N/A
His-tagged $McyB_1$-(A-PCP)	This paper	N/A
His-tagged $McyB_1$-(C-A^{R871A})	This paper	N/A
His-tagged $McyB_1$-(C-A^{R874A})	This paper	N/A
His-tagged $McyB_1$-(C^{E359A}-A)	This paper	N/A
His-tagged $McyB_1$-(C^{M213A}-A)	This paper	N/A
His-tagged $McyB_1$-(C^{R871C}-A)	This paper	N/A
His-tagged $McyB_1$-(C^{M213C}-A)	This paper	N/A
His-tagged $McyB_1$-(C^{M213C}-A^{R871C})	This paper	N/A
His-tagged $McyA_2$-(C-A-PCP)	This paper	N/A
His-tagged $McyA_2$-(C-A)	This paper	N/A
His-tagged $McyA_2$-(A)	This paper	N/A
His-tagged $McyA_2$-(A-PCP)	This paper	N/A
His-tagged $McyA_2$-(C-A^{R2129A}-PCP)	This paper	N/A
His-tagged $McyA_2$-(C-A^{R2132A}-PCP)	This paper	N/A
His-tagged $McyA_2$-(C^{D1481A}-A-PCP)	This paper	N/A
His-tagged $McyA_2$-(C^{S1611A}-A-PCP)	This paper	N/A
His-tagged $McyA_2$-(C-A^{R2129C}-PCP)	This paper	N/A
His-tagged $McyA_2$-(C^{S1611C}-A-PCP)	This paper	N/A
His-tagged $McyA_2$-(C^{S1611C}-A^{R2129C}-PCP)	This paper	N/A
His-tagged McyC-(C-A-PCP)	This paper	N/A
His-tagged McyC-(C-A)	This paper	N/A
His-tagged McyC-(A)	This paper	N/A
His-tagged McyC-(C-A^{R880A}-PCP)	This paper	N/A
His-tagged McyC-(C-A^{R877A}-PCP)	This paper	N/A
His-tagged EntF-(C-A)	This paper	N/A
His-tagged EntF-(A)	This paper	N/A
His-tagged EntF-(C-A^{R863A})	This paper	N/A
His-tagged EntF-(C-A^{R866A})	This paper	N/A
His-tagged SrfA-C-(C-A-PCP)	This paper	N/A
His-tagged SrfA-C-(A)	This paper	N/A
His-tagged SrfA-C-(C-A^{R867A}-PCP)	This paper	N/A
His-tagged SrfA-C-(C-A^{R870A}-PCP)	This paper	N/A
Deposited data		
Crystal structure of $McyB_1$-(C-A)	This paper	PDB: 8HLK
Crystal structure of $McyA_2$-(C-A-PCP)	This paper	PDB: 8JBR

Structure
Article

CellPress

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
Please refer to Table S1	This paper	N/A
Oligonucleotides		
Please refer to Table S2	This paper	N/A

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by Xiaoli Zeng (zengxl@ihb. ac.cn).

Materials availability
Most of the materials used in this study are publicly or commercially available. All unique and stable reagents generated in this study are available from the lead contact without restriction, upon reasonable request.

Data and code availability
The accession numbers for the structures reported in this paper are PDB: 8HLK, 8JBR.

All figures showing a structure were prepared with PyMOL (https://pymol.org/2/). This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available through contacting.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Microbe strains
E. coli DH5α competent cells were used for cloning and plasmid preparation. *E. coli* BL21(DE3) competent cells were used for protein expression. All cells are commercially available and were cultured as described in the method details section.

METHOD DETAILS

Plasmids construction
The plasmids, used for protein expression and purification, were constructed based on the pET28a expression vector, which contains the sequence encoding a 6XHis tag. DNA fragments encoding the single A domain, the C-A and A-PCP didomains, or the C-A-PCP tri-domain polypeptides of McyA$_2$, McyB$_1$ and McyC were PCR amplified, respectively, from the *Microcystis* genomic DNA (GenBank: AM778952.1). DNA fragments encoding the single A domain, or C-A-PCP tri-domain of EntF were PCR amplified, respectively, from *E. coli* JM109 (The primers were designed according to the K-12 genome sequence, GenBank: NC_000913.3). DNA fragments encoding the single A domain or C-A didomain of SrfA-C were amplified, respectively, by PCR from genomic DNA of the surfactin producer strain *B. subtilis* ATCC 21332. The amplified DNA fragments were cloned into pET28a, transformed into *E. coli* DH5α, and confirmed by sequencing. To construct the plasmids containing genes with point mutation(s), quick-change mutagenesis was carried out using the plasmid containing the corresponding WT genes as template. Plasmids, and oligonucleotides used in this study are listed in Tables S1 and S2, respectively.

Protein expression and purification
The plasmids for recombinant protein expression were transformed into *E. coli* BL21 (DE3) and the strains were grown in Luria-Bertani (LB) medium containing kanamycin (50 mg/ml) with shaking at 37°C until OD$_{600}$ of 0.4–0.6. Protein expression was induced overnight with addition of 0.5 mM isopropyl-β-d-thiogalactoside (IPTG) at 16°C. Cells were collected by centrifugation at 9000 × g for 10 min. Selenomethionine (SeMet)-labeled proteins were overexpressed in *E. coli* BL21 (DE3) (Novagen). Transformed cells were first cultured in LB medium at 37°C overnight, then harvested and washed twice with M9 medium.[51] Then cells were cultured in SeMet medium (M9 medium with 50 mg L^{-1} SeMet and other essential amino acids) to OD$_{600}$ of 0.6–0.8.

The native and SeMet labeled proteins was produced as described previously, with minor modifications.[52] Cell pellets were resuspended in lysis buffer (200 mM NaCl, 10% glycerin, 20 mM Tris-HCl pH 8.0, 0.2 mM TCEP). Cells were lysed using French press (JINBO,Inc), followed by centrifugation at 12000 × g for 40 min at 4°C. The initial purification was achieved with a Ni-NTA column. The supernatant was mixed with 10% (v/v) Ni-NTA agarose beads (Qiagen) preequilibrated with the lysis buffer for 40 min at 4°C. The beads were then collected by filtration and washed with the lysis buffer and the washing buffer (200 mM NaCl, 10% glycerin, 20 mM Tris-HCl pH 8.0, 0.2 mM TCEP, 20 mM imidazole). Protein was eluted using elution buffer (200 mM NaCl, 10% glycerin, 20 mM Tris-HCl pH 8.0, 0.2 mM TCEP, 250 mM imidazole). The eluted samples were concentrated (Amicon Ultracel-10K, Millipore) and further

purified by size-exclusion chromatography (SEC) with a HiLoad 26/600 Superdex 200 column (GE Healthcare). The column was equilibrated with the SEC buffer (200 mM NaCl, 10% glycerol, 20 mM Tris-HCl pH 8.0, 0.2 mM TCEP). The purity of each fraction was determined by SDS-PAGE. Fractions containing the pure recombinant protein were pooled, concentrated and stored at $-80^{\circ}C$ for further use.

Crystallization

Before setting up for crystallization, purified native proteins or SeMet-labelled $McyA_2$-(C-A-PCP) were phosphopantetheinylated by incubation with the phosphopantetheinyl transferase Sfp (10 nM), in the presence of 12.5 mM $MgCl_2$ and 1 mM CoA for 60 min at $20^{\circ}C$. Then, the protein sample was passed over a HiLoad 26/600 Superdex 200 column (GE Healthcare) pre-equilibrated with the equilibration buffer (100 mM NaCl, 20 mM Tris-HCl pH 8.0, 0.2 mM TCEP) to remove the Sfp protein. The eluted holo-$McyA_2$-(C-A-PCP) protein was pooled and concentrated to 15 mg/ml by ultrafiltration (Amicon Ultracel-10K, Millipore), then, mixed with 2 mM ATP, 2 mM L-Ala and 3 mM $MgCl_2$ at $4^{\circ}C$ for 1 h. The crystallization conditions were initially identified from a sparse matrix screen at $20^{\circ}C$. Final crystals for SeMet-labelled holo-$McyA_2$-(C-A-PCP) were obtained against the reservoir solution containing 0.18 M lithium sulfate monohydrate,0.1 M Tris-Cl pH 8.0,17% w/v polyethylene glycol 3350 using the hanging drop vapor diffusion method at $20^{\circ}C$, with a drop of 1 μL protein solution mixed with 1 μL reservoir.

For $McyB_1$-(C-A), the purified native protein was first concentrated to 8 mg/ml by ultrafiltration (Amicon Ultracel-10K, Millipore), then mixed with 2 mM ATP, 2 mM L-Leu and 3 mM $MgCl_2$ at $4^{\circ}C$ for 1 h. The initial crystallization conditions for the $McyB_1$-(C-A) were identified from a sparse matrix screen at $20^{\circ}C$. Final high-quality crystals were grown against the reservoir solution containing 3.5 M sodium formate, 5% glycerol, 80 mM sodium chloride, and 50 mM Tris-HCl pH 8.0 using hanging drop vapor diffusion at $20^{\circ}C$, with a drop of 1 μL protein solution mixed with 1 μL reservoir.

Structure determination of $McyA_2$-(C-A-PCP) and $McyB_1$-(C-A)

SeMet-labelled crystals of holo-$McyA_2$-(C-A-PCP) were flash-cooled with 30% glycerol as cryoprotectant in liquid nitrogen. X-ray diffraction data was collected at 100 K on an ADSC Q315r CCD detector at beamline BL19U of Shanghai Synchrotron Radiation Facility (SSRF). Diffraction data was indexed, integrated and scaled using *HKL2000*[53] in space group *P3121*. The structure was solved by molecular replacement with *Molrep* in CCP4i.[54] Initially, we solved the structure of the C domain using the AlphaFold[55] predicted model of $McyA_2$ (residue I1267-S1710) as an initial search model. Afterwards, via fixing the solution of the C domain, the MR was performed using the A domain (residue E1711-S2224, predicted by AlphaFold) of $McyA_2$ as a subsequent search model. Finally, the PCP domain was solved by MR using the model of AB3403 (PDB:4ZXH, residues A971-E1045) via fixing the C and A domains. The maximum-likelihood method implemented in *REFMAC5*[56] as part of the *CCP4* program suite[54] and *Coot*[57] were used interactively to build the linkage among different domains manually, and the structure was automatically refined in the program Phenix.[58] The ligands and waters were manually built in the F_O-F_C maps in Coot to produce the final model. The ATP molecule of $McyA_2$-(C-A-PCP) was manually adjusted in Coot and automatically refined in REFMAC5 of CCP4. The final model was evaluated with *PROCHECK*[59] and *MolProbity*.[60]

Crystals of $McyB_1$-(C-A) were flash-cooled with 20% glycerol as cryoprotectant in liquid nitrogen. In total, 1157 images of diffraction data were collected using a Rigaku RUH-3R rotating copper-anode source equipped with a PILATUS 200K detector (DECTRIS, Switzerland) at 100 K. Diffraction data was indexed, merged, and scaled using *iMOSFLM*[61] in space group *I 41* (2.70 Å). Structure determination was performed in *PHENIX* using molecular replacement and the structure of AB3403 (PDB: 4ZXH, residues M1-E964) was used as the search model. The subsequent refinement process is similar to holo-$McyA_2$-(C-A-PCP). Crystallographic parameters and data-collection statistics for both structures are listed in Table 1.

Continuous hydroxylamine release assay

The adenylation activity of A domains in different proteins was determined by the continuous hydroxylamine release assay (Figure S2). This method is similar to the hydroxylamine release assay described previously, with minor modifications.[62] Briefly, 300 μL freshly prepared solution A (0.04 U inorganic phosphatase, 150 mM hydroxylamine, 5 mM ATP, 5 mM $MgCl_2$,100 mM Tris-HCl pH 8.0) was first mixed with 6 μL 100 mM amino acid solution at $25^{\circ}C$ for 5 min. The adenylation reaction was initiated by addition of 6 μL (1 μM/L) protein in 1.5-ml tubes at $25^{\circ}C$ for 10 min. Then, 300 μl dyeing solution (Phosphate colorimetric assay kit, Sigma-Aldrich) was transferred into the reaction tubes and incubated at $25^{\circ}C$ for 30 min. After incubation, 200 μl reaction solution was transferred into a 96-well plate and the absorbance at 650 nm measured by using Spectramax M5e.

Bioinformatic analysis

The sequence identity or similarity of NPRS homologs with C-A didomains was calculated using Clustal Omega software. Multiple sequence alignment was presented by ESPript 3.0 with 5 representative sequences of C-A didomains from *Microcystis aeruginosa* PCC 7806, *Escherichia coli* JM109 and *Bacillus subtilis* ATCC 21332. The sequence logo of the A8 catalytic motif in NRPS elongation or termination modules are made by the WebLogo server with 79 representative modules (Table S3). The sequence logo of the A8 catalytic motif in initiation module or standalone A domain of different NRPSs are made by the WebLogo server with 21 representative NRPS module (Table S3). These NRPSs are from different bacteria, including *Microcystis aeruginosa* PCC 7806, *Escherichia coli* JM109, *Bacillus subtilis* ATCC 21332, *Acinetobacter baumannii* ATCC 17978, *Brevibacillus parabrevis* ATCC 8185, *Burkholderia diffusa*, *Streptomyces atroolivaceus*, *Pseudomonas protegens* Pf-5, *Pantoea agglomerans*, *Bacillus velezensis* YAU B9601-Y2,

Structure
Article

Streptomyces rapamycinicus NRRL 5491, *Amycolatopsis rubida*, *Brevibacillus* sp. CF 112, *Brevibacillus* sp. BC 25, *Amycolatopsis orientalis*, *Bacillus licheniformis*, *Nodularia spumigena* UHCC 0039, *Bacillus brevis* ATCC 999, *Streptomyces roseosporus* NRRL 11379, *Mycobacterium smegmatis* ATCC 700084, *Pseudomonas aeruginosa* PAO1 and *Streptomyces halstedii*.

QUANTIFICATION AND STATISTICAL ANALYSIS

Two biological replicates were carried out for each experimental condition unless otherwise indicated. Statistical tests employed are indicated in the text and the corresponding figure legends. Analysis was performed in Excel (Microsoft) or Origin software. No statistical methods were used to predetermine sample size.