Current Research in Structural Biology 2 (2020) 14–24

Contents lists available at ScienceDirect

Current Research in Structural Biology

journal homepage: www.journals.elsevier.com/current-research-in-structural-biology

Structures of teixobactin-producing nonribosomal peptide synthetase condensation and adenylation domains☆

Kemin Tan [a,b], Min Zhou [a,c], Robert P. Jedrzejczak [a,b], Ruiying Wu [a,b], Raul A. Higuera [d], Dominika Borek [e], Gyorgy Babnigg [a,c], Andrzej Joachimiak [a,b,f,*]

[a] Center for Structural Genomics of Infectious Diseases, University of Chicago, 5735 South Ellis Avenue, Chicago, IL 60637, USA
[b] Structural Biology Center, X-ray Science Division, Advanced Photon Source, Argonne National Laboratory, 9700 S. Cass Ave. Lemont, IL 60439, USA
[c] Biosciences, Argonne National Laboratory, Argonne, IL 60439, USA
[d] BUILDing SCHOLARS, The University of Texas at El Paso, 500 W. University Ave, El Paso, TX 79968, USA
[e] Department of Biophysics, UT Southwestern Medical Center, Dallas, TX 75390, USA
[f] Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60367, USA

ARTICLE INFO

Keywords:
Nonribosomal peptide synthetase
Txo1
Txo2
Teixobactin
Condensation domain
Adenylation domain
Conformational change
α-helix regulation
Mg^{2+}-binding

ABSTRACT

The recently discovered antibiotic teixobactin is produced by uncultured soil bacteria. The antibiotic inhibits cell wall synthesis of Gram-positive bacteria by binding to precursors of cell wall building blocks, and therefore it is thought to be less vulnerable to development of resistance. Teixobactin is synthesized by two nonribosomal peptide synthetases (NRPSs), encoded by *txo1* and *txo2* genes. Like other NRPSs, the Txo1 and Txo2 synthetases are large, multifunctional, and comprised of several modules. Each module is responsible for catalysis of a distinct step of teixobactin synthesis and contains specific functional units, commonly including a condensation (C) domain, an adenylation (A) domain, and a peptidyl carrier protein (PCP) domain. Here we report the structures of the C-A bidomains of the two L-Ser condensing modules, from Txo1 and Txo2, respectively. In the structure of the C domain of the L-Ser subunit of Txo1, a large conformational change is observed, featuring an outward swing of its N-terminal α-helix. This repositioning, if functionally validated, provides the necessary conformational change for the condensation reaction in C domain, and likely represents a regulatory mechanism. In an A_{core} subdomain, a well-coordinated Mg^{2+} cation is observed, which is required in the adenylation reaction. The Mg^{2+}-binding site is defined by a largely conserved amino acid sequence motif and is coordinated by the α-phosphate group of AMP (or ATP) when present, providing some structural evidence for the role of the metal cation in the catalysis of A domain.

1. Introduction

Nonribosomal peptide synthetases (NRPSs) are large multifunctional enzymes, which consist of multiple modules that bind and catalyze in an assembly-line fashion the addition of amino acid monomers to create a variety of complex secondary metabolites (Marahiel et al., 1997; Fischbach & Walsh, 2006; Weissman, 2015; Miller & Gulick, 2016; Payne et al., 2016). The nonribosomal peptides produced in bacteria are of great interest to the pharmaceutical industry because of their antibacterial, antiviral, and anticancer properties (Walsh, 2015; Gulick, 2017; Agrawal

Abbreviations: NRPS, Nonribosomal peptide synthetase; C domain, Condensation domain; C_{Nterm} subdomain, N-terminal subdomain of C domain; C_{Cterm} subdomain, C-terminal subdomain of C domain; A domain, Adenylation domain; A_{core} subdomain, Large N-terminal subdomain of A domain; A_{sub} subdomain, Small C-terminal subdomain of A domain; PCP domain, Peptidyl carrier domain; COM^D domain, Donor communication-mediating domain; COM^A domain, Acceptor communication-mediating domain; SAD, Single wavelength diffraction; RMSD, Root-mean-square deviation; SSM, Secondary-structure matching; MES, 2- morpholinoethane sulfonic acid.

* Corresponding author. University of Chicago, Department of Biochemistry and Molecular Biology, Chicago IL 60367, United States. Tel.: 630 252 392.
E-mail address: andrzejj@anl.gov (A. Joachimiak).

https://doi.org/10.1016/j.crstbi.2020.01.002
Received 5 July 2019; Received in revised form 9 January 2020; Accepted 13 January 2020

et al., 2017). The newly discovered nonribosomal compound teixobactin, produced by an uncultured soil beta-proteobacterium, *Eleftheria terrae*, represents the first of a new class of antibiotics against Gram-positive organisms (Ling et al., 2015; Girt et al., 2018) that kills important pathogens, including methicillin resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant *Enterococci* (VRE), penicillin-resistant *Streptococcus pneumoniae* (PRSP), and *Mycobacterium tuberculosis* without detectable resistance, although it is expected that resistance will eventually develop in clinical settings. Teixobactin also shows some activity against *Clostridium difficile*, which is known to cause nosocomial diarrhea, and *Bacillus anthracis*. Teixobactin is thought to bind highly conserved prenyl-pyrophosphate-saccharide regions of lipid II and related membrane-bound cell wall precursors (Ling et al., 2015; Homma et al., 2016).

Teixobactin is a cyclic peptide comprised L- and D-amino acids (N–Me-D-Phe$_1$, L-Ile$_2$, L-Ser$_3$, D-Gln$_4$, D-allo-Ile$_5$, L-Ile$_6$, L-Ser$_7$, D-Thr$_8$, L-Ala$_9$ and L-allo-enduracitidine$_{10}$) synthesized by two mega NRPSs, Txo1 (6422 a.a.) and Txo2 (6375 a.a.) (Fig. 1A). Txo1 and Txo2 contain six and five functional modules, respectively. Like other NRPSs, each module contains a C domain, an A domain, and a PCP domain (Challis & Naismith, 2004). In addition, Txo1 contains a methylation (MT) domain and Txo2 has two thioesterase (TE) domains (Mandalapu et al., 2018). Individual domains share structural similarity with other NRPS domains but show varied degrees of sequence divergence. The COM^D domain located at the C-terminus of Txo1 and the COM^A domain located at the N-terminus of Txo2, which are essential for Txo1 and Txo2 to recognize and associate with each other to ensure the assembly of defined product (Hahn & Stachelhaus, 2006; Dehling et al., 2016), await further investigation.

The A domain is responsible for selective substrate binding, substrate adenylation, and transfer of the adenylated substrate to the thiol of the pantetheine cofactor of the PCP domain (Mootz & Marahiel, 1997; Conti et al., 1997; Stachelhaus et al., 1999; Bloudoff et al., 2016; Drake et al., 2016). The adenylation reaction involves ATP and Mg^{2+}. The binding sites of ATP and AMP (the product of adenylate- and thioester-forming reactions) have been extensively studied in several NRPSs (Conti et al., 1997; May et al., 2002; Yonus et al., 2008). A magnesium ion is required in the adenylation reaction (Airas, 2007; Schmelz & Naismith, 2009). However, in currently available A domain structures, Mg^{2+}-binding sites are poorly defined with weak electron densities for the metal ion and an incomplete coordination sphere (Conti et al., 1997; Drake et al., 2016; Gulick et al., 2003). In addition, it has been proposed that A domain undergoes a large conformational rearrangement during the catalytic reaction, with a particularly significant reorientation between its large (A_{core}) and small (A_{sub}) subdomains (Yonus et al., 2008; Strieker et al., 2010; Gulick, 2016; Mitchell et al., 2012).

There is less information about the conformational changes of C domain itself and other members of the C domain superfamily (Bloudoff et al., 2016; Bloudoff et al., 2013; Bloudoff & Schmeing, 2017; Chen et al., 2016; Dowling et al., 2016; Zhang et al., 2016). The catalytic site of C domain is located in a narrow tunnel between its two subdomains, the N-terminal (C_{Nterm}) and the C-terminal (C_{Cterm}) halves (Bloudoff & Schmeing, 2017). It is believed that a "latch" above the catalytic site needs to open for substrates to access key active site residues, including the second His residue within the highly conserved sequence motif HHxxxDG (Samel et al., 2007). Thus far, very limited relative orientation changes between the two subdomains from different NRPS C domains have been reported (Bloudoff et al., 2013; Bloudoff & Schmeing, 2017), and none of these conformations are associated with the opening of the "latch" above the catalytic site. According to a modeling of the catalytic mechanism (Bloudoff et al., 2013; Bloudoff & Schmeing, 2017; Samel et al., 2007), upstream (donor) substrate and downstream (acceptor) substrate approach an active site from the two opposite directions of the narrow tunnel between the two subdomains (Fig. 1B). Available structures with small acceptor substrate molecules in the tunnel suggest the C domain can accommodate small substrates without a major conformational change of the domain (Bloudoff et al., 2016; Drake et al., 2016). On the other hand, the mechanism by which donor PCP positions its attached substrate inside the tunnel still remains unknown. As peptides grow larger, especially during the later stages of NRPS reactions, the tunnel seems to be too narrow to accommodate a growing peptide attached to the donor PCP domain at one end of the tunnel. This also applies to the elongated peptide after condensation, attached to the downstream acceptor PCP domain. It is supposed to move through the

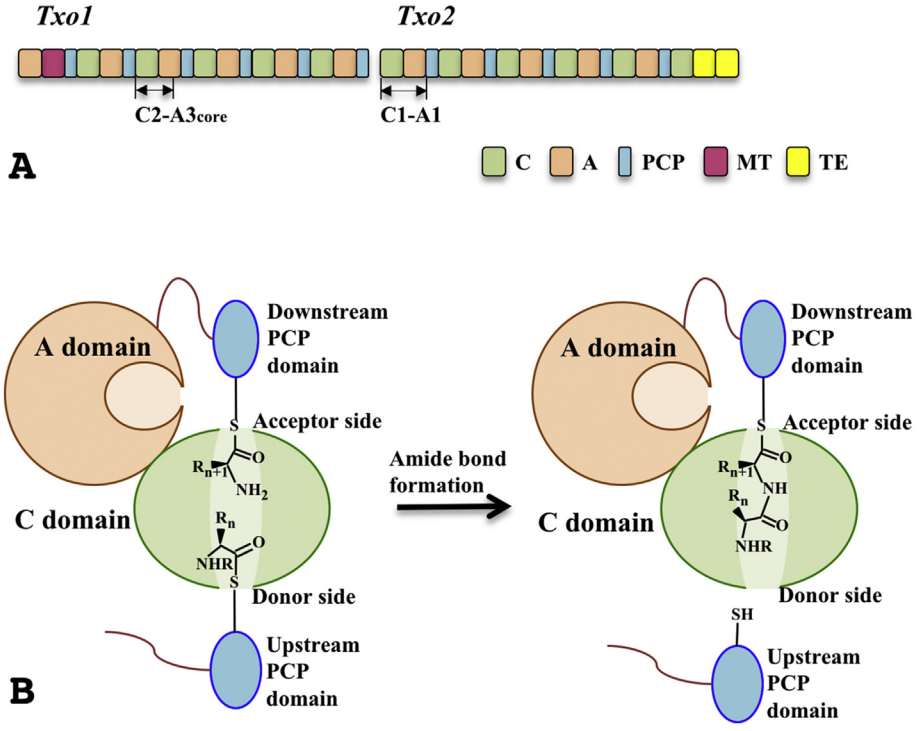

Fig. 1. Antibiotic teixobactin synthesizing genes, *Txo1* and *Txo2* and a proposed catalytic model for condensation (C) domain. (A) Domain arrangement of non-ribosomal peptide synthetases Txo1 and Txo2 (Ling et al., 2015). C domains, adenylation (A) domains, peptidyl carrier protein (PCP) domains, methylation (MT) domain and thioesterase (TE) domains are colored differently as indicated. The COM^D domain located at the C-terminus of Txo1 and the COM^A domain located at the N-terminus of Txo2 were not shown in the diagram. The two protein expression constructs, Txo1_C1-A3$_{core}$ and Txo2_C1-A1, are mapped to the domain's arrangement diagram. (B) A proposed model of amide bond formation within a C domain. $R_{n+1}ONH_2$ represents the cognate amino acid of the current subunit. R_nONHR represents the intermediate product from upstream subunit.

tunnel for its release at the other end (Fig. 1B). The protein conformational changes during catalytic cycle cannot be accurately modelled because only a close conformation of the C domain has been reported.

It is believed that the teixobactin is synthesized in stepwise fashion. To initiate synthesis, the first module of Txo1 (Txo1_A1_MT1_PCP1) produces a *N*-methylated phenylalanine and attaches it to the first PCP domain (Ling et al., 2015). The second module adds L-Ile. The third module of Txo1 is a serine condensing (Ser)-module adding L-Ser_3, including the second C domain (Txo1_C2), the third A domain (Txo1_A3) and the third PCP domain (Txo1_PCP3) (Ling et al., 2015). There is also a Ser-module in Txo2 responsible for addition L-Ser_7. This is the first module in Txo2 and, like Txo1, it includes a C domain (Txo2_C1), an A domain (Txo2_A1) and a PCP domain (Txo2_PCP1) (Ling et al., 2015). The A domains of the two Ser modules, Txo1_A3 and Txo2_A1, are essentially alike with a sequence identity of 99.6% (SI Fig. 1). This is not surprising because the role of both A domains is the same, i.e., they need to recognize and activate serine with high reliability. The C domains of the two Ser modules, Txo1_C2 and Txo2_C1, have a sequence identity of 80.1% (SI Fig. 2). Although, in both cases the adenylated serine is attached to the acceptor PCP domain, the peptides attached to the donor PCP domains are in different stages of elongation (3rd and 7th respectively).

In our studies of Txo1 and Txo2, which are community nominated targets of the Center for Structural Genomics of Infectious Diseases (CSGID), multiple constructs of the two Ser-modules were designed and expressed. Diffraction quality crystals were obtained for constructs Txo2_C1-A1 (residues D45-L1005) and Txo1_C2-$A3_{core}$ (residues P2140-G3009). In the Txo1_C2-$A3_{core}$ construct, the C-terminal A_{sub} subdomain (residues L3010-V3109) is truncated. Their structures were subsequently determined as described in Materials and Methods. In order to explore the ligand binding properties of an A domain, we also obtained co-crystals of Txo1_C2-$A3_{core}$ with AMP and Mg^{2+} and determined their structures. Here we focus on an open conformation of a C domain (Txo1_C2), as well as a well-defined Mg^{2+}-binding site in an A domain (Txo1_$A3_{core}$).

2. Materials and methods

2.1. Cloning and protein expression

The gene constructs for the Txo1 (AJF34463) Ser module C-A bidomain of (Txo1_C2-$A3_{core}$, residues P2140-G3009) and the Txo2 (AJF34463) full length Ser module C-A bidomain of (Txo2_C1-A1, residues D45-L1005) from *Eleftheria terrae* (genomic DNA kindly provided by Dr. Kim Lewis) were PCR amplified. Purified PCR products were treated with T4 DNA polymerase in the presence of dCTP (Eschenfeldt et al., 2010) according to vendor specification (New England Biolabs, Ipswich, Massachusetts, USA) and cloned into expression vector pMCSG56 using ligation independent cloning (Aslanidis & de Jong, 1990; Eschenfeldt et al., 2009). Vector compatible primers for amplification of DNA fragments coding for the subunits were designed using the online tool (https://bioinformatics.anl.gov/targets/public_tools.aspx) (Yoon et al., 2002). The expression vectors were then transformed into *E. coli* pGro7-K cells. The starter cultures were grown at 37 °C overnight in 500 mL bottles containing 25 mL of modified M9 media (Stols et al., 2004). The cultures were then transferred to a 2 L bottle containing 1 L of the same media for growth to an OD_{600} of 1.4. They were cooled down to 18 °C for 15 min before inhibitory amino acids (25 mg each of L-valine, L-isoleucine, L-leucine, L-lysine, L-threonine, L-phenylalanine), 60 mg selenomethionine (Medicillin, Inc., for SeMet-labeled protein only), and 1 mM IPTG were added to the culture. Cultures were then grown overnight at 18 °C and harvested the next morning. Cells were harvested from 2 L of culture by centrifugation and the cell pellet was re-suspended in 60 mL of lysis buffer containing 50 mM HEPES pH 8.0, 500 mM NaCl, 5% (v/v) glycerol, 10 mM imidazole, 10 mM β-ME, and 1 protease inhibitor cocktail tablet (Complete, Roche). Re-suspended cells were stored at −80 °C before processing.

2.2. Protein purification

Frozen cells were thawed and sonicated on ice with one protease inhibitor cocktail tablet (Complete, Roche) per liter of cell culture. Lysates were sonicated for 5 min by means of a program or with manual control using a 4-s on and 20-s off mode and power/voltage settings according to the manufacturer's instructions and centrifuged at 30,000 g for 60 min, followed by syringe filtration (0.45 μm). Clarified lysate was loaded onto a 5-mL nickel HisTrap HP column (GE Healthcare Life Sciences) and the His_6-tagged protein was released with elution buffer (500 mM NaCl, 5% glycerol, 50 mM HEPES, pH 8.0, 250 mM imidazole, and 10 mM 2-mercaptoethanol). This step was followed by size exclusion chromatography on a Superdex 200 HiLoad 26/60 column (GE Healthcare Life Sciences) in crystallization buffer (250 mM NaCl, 20 mM HEPES pH 8.0, 2 mM DTT). All these steps were performed on an ÄKTAxpress system (GE Healthcare Life Sciences). The fusion tag was removed by a 48 h digest with recombinant His_7-tagged Tobacco Etch Virus (TEV) protease. The digestion was verified by SDS-PAGE and showed complete removal of the His_6-tag. Nickel affinity chromatography was used to remove the His_6 tag, uncut protein, and His_7-tagged TEV protease (Blommel & Fox, 2007). The proteins were then concentrated and buffer-exchanged with a crystallization buffer via ultrafiltration.

2.3. Protein crystallization

Both wild-type and SeMet-labeled proteins were screened for crystallization conditions with the help of a Mosquito nanoliter liquid handler (TTP LabTech) using the sitting drop vapor diffusion technique in 96-well CrystalQuick plates (Greiner). For each condition, 0.5 μl of protein (at ~40 mg/ml for Txo1_C2-$A3_{core}$ and ~90 mg/ml for Txo2_C1-A1) and 0.5 μl of crystallization formulation were mixed; the mixture was equilibrated against 150 μl of the crystallization solution in each reservoir well. The MCSG-1-4 (Microlytic) crystallization screens were used for the screening at temperature of 16 °C. Crystals appeared under multiple conditions and diffraction quality crystals leading to structure determination were from several conditions. Examples of crystallization conditions for SeMet-labeled and native Txo1_C2-$A3_{core}$ were: A) 0.1 M MES:NaOH, pH 6.0 and 1.26 M $(NH_4)_2SO_4$; B) 1.6 M $MgCl_2$ and 0.1 M MES:NaOH, pH 6.5; and C) 0.1 M HEPES pH 7.5 and 1.26 M $(NH_4)_2SO_4$. Crystals from conditions (B) and (C) were also used for AMP soaking by transferring crystals to their soaking buffers (mother liquid plus 25 mM AMP) for 30 min. The best crystallization condition for wild type Txo2_C1-A1 was 0.2 M sodium citrate, 0.1 M Tris:HCl pH 8.5, 15% (v/v) PEG400. Prior to X-ray diffraction data collection, all crystals were treated with a cryoprotectant solution (mother liquid plus 30% glycerol) for seconds and cryocooled directly in liquid nitrogen.

2.4. X-ray diffraction and structure determination

Single-wavelength X-ray diffraction data were collected near the selenium absorption peak (12.66 keV) at 100 K from crystals of SeMet-labeled Txo1_C2-$A3_{core}$, native crystals of Txo1_C2-$A3_{core}$ and Txo2_C1-A1. The data were obtained at the 19-ID beamline of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory using the program SBCcollect (Rosenbaum et al., 2006). The intensities of each data set were integrated and scaled with the HKL3000 program suite (Minor et al., 2006) (Table 1). The structure of SeMet-labeled Txo1_C2-$A3_{core}$ was determined first using single-wavelength anomalous dispersion (SAD) method. There was one Txo1_C2-$A3_{core}$ monomer in one asymmetric unit. Selenium sites were first located using the program SHELXD (Schneider & Sheldrick, 2002) and they were used for phasing with the program MLPHARE (Winn et al., 2011). After density modification, partial models were built in cycles of automatic model building using HKL Builder. All of the above programs are integrated within the program suite HKL3000. The final

Table 1
Data collection and refinement statistics.

Data collection	Txo2_C1_A1	Txo1_C2_A3$_{core}$	Txo1_C2_A3$_{core}$ + AMP	Txo1_C2_A3$_{core}$ + Mg^{2+}	Txo1_C2_A3$_{core}$ + AMP + Mg^{2+}
Space group	$C222_1$	$C2$	$C2$	$C2$	$C2$
Unit Cell Dimensions					
a, b, c (Å)	110.9, 399.9, 144.1	153.0, 90.75, 98.22	154.7, 90.86, 98.18	154.5, 90.66, 98.80	154.2, 90.83, 98.46
α, β, γ (°)	90, 90, 90	90, 106.0, 90	90, 106.7, 90	90, 106.0, 90	90, 106.1, 90
Protein MW Da (# of residues)	105,717.9 (961)[a]	97,306.9 (880)[a]	97,306.9 (880)[a]	97,306.9 (880)[a]	97,306.9 (880)[a]
Wavelength (Å)	0.9717	0.9792	0.9792	0.9792	0.9792
Resolution (Å)	47.2–2.95	43.1–2.10	47.5–2.18	47.5–2.15	46.1–2.10
Number of Unique reflections	66,302	75,359	65,631	70,542	73,508
Completeness (%)	97.8 (99.1)[b]	98.9 (98.0)[c]	96.8 (82.0)[d]	98.6 (92.3)[e]	96.3 (82.2)[f]
Redundancy	3.4 (3.3)[b]	3.4 (3.3)[c]	3.7 (3.3)[d]	4.3 (3.3)[e]	3.6 (3.2)[f]
R_{merge}	0.111 (0.837)[b]	0.060 (0.766)[c]	0.060 (0.618)[d]	0.082 (0.785)[e]	0.053 (0.652)[f]
$CC_{1/2}$	0.901 (0.710)	0.955 (0.628)	0.976 (0.714)	0.983 (0.699)	0.997 (0.692)
I/σ(I)	10.6 (1.3)[b]	20.4 (1.2)[c]	19.2 (1.7)[d]	28.9 (1.3)[d]	22.6 (1.1)[f]
Solvent content (%)	67.3	63.5	64.1	65.7	65.4
Wilson B-factors ($Å^2$)	54.1	22.7	29.1	49.5	49.0
Phasing					
Resolution (Å)	47.2–3.00	43.1–3.00	47.5–3.00	47.5–2.15	46.1–3.00
CullR_ano (%)				0.85	
FOM before DM				0.15	
Correlation coefficient[g] coefficient	0.44	0.68	0.67		0.66
Refinement					
Resolution	47.2–2.95	43.1–2.10	47.5–2.18	47.5–2.15	46.1–2.10
No. reflections/Test set/Test (work/test)	66,102/3241	66,186/3285	57,192/2810	70,279/3391	73,414/3670
R_{work}/R_{free}	0.226/0.267	0.186/0.214	0.192/0.218	0.201/0.231	0.208/0.236
No.of atoms					
Protein/Water/Others	14,876/77/104	6527/406/107	6363/277/49	6209/117/237	6431/157/129
B-factors ($Å^2$)					
Protein/Water/Others	73.1/43.2/73.2	54.9/37.2/88.3	61.3/39.8/57.5	85.6/57.2/111.8	82.5/53.1/123.4
R.m.s deviation					
Bond length (Å)	0.002	0.003	0.002	0.002	0.002
Bond angle (°)	0.458	0.559	0.487	0.477	0.485
Ramachandran Plot (%)					
Favored regions	96.96	95.63	96.24	97.43	96.65
Outliers	0.37	0.71	0.24	0.37	0.60
PDB ID	**6P1J**	**6OYF**	**6OZV**	**6P3I**	**6P4U**

[a] Not including three N-terminal vector-derived residues, SNA.
[b] Last resolution bin, 2.95-3.04 Å.
[c] Last resolution bin, 2.10-2.14 Å.
[d] Last resolution bin, 2.18-2.22 Å.
[e] Last resolution bin, 2.15-2.2.19 Å.
[f] Last resolution bin, 2.10-2.14 Å.
[g] Molecular replacement.

Txo1_C2-A3$_{core}$ model was completed after alternative cycles of manual building using the program COOT (Emsley & Cowtan, 2004) and refinements using the programs of Phenix.Refine (Afonine et al., 2012) (Table 1). The individual domains (Txo1_C2 and Txo1_A3$_{core}$) resolved in the Txo1_C2-A3$_{core}$ structure were the best search templates in the structure determination of Txo2_C1-A1 using molecular replacement method with MolRep (Vagin & Teplyakov, 2010). The structures of Txo1_C2-A3$_{core}$ crystals soaked with AMP were determined using the Txo1_C2-A3$_{core}$ bidomain as search template. The subsequent refinements of these structures were completed in a similar procedure as mentioned above for the Txo1_C2-A3$_{core}$ structure. Structure validation was performed using the program MolProbity (Williams et al., 2018). The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) (Table 1).

3. Results

3.1. Txo2_C1-A1 structure

As mentioned earlier the Txo2 starts with a serine module, including Txo2_C1, _A1 and _PCP1 domains. We have determined the structure of the Txo2_C1-A1 bidomain construct at 2.95 Å, using molecular replacement method (Table 1). In the crystal of Txo2_C1-A1, there are two monomers (A and B) in one asymmetric unit. They are packed in a pseudo 2-fold symmetry without any recognizable dimerization interface between each other. The structures of the two monomers are nearly identical with a RMSD value of 0.42 Å when they are superimposed by using secondary-structure matching (SSM) (Krissinel & Henrick, 2004). In the following description and discussion, only monomer A will be used.

The overall structure of C1 domain resembles other condensation domains available at the PDB (Samel et al., 2007; Keating et al., 2002; Tanovic et al., 2008). Its two structurally similar subdomains, N-terminal subdomain (Txo2_C1$_{Nterm}$) and C-terminal subdomain (Txo2_C1$_{Cterm}$) have a chloramphenicol acetyltransferase (CAT) fold, which typically consists of a six-stranded β-sheet and five-helices packed against one face of the sheet (Leslie et al., 1988) (Fig. 2A and B). The two subdomains are in a V-shape arrangement with their linker region forming a "floor" (Fig. 2B). There are also two crossovers from Txo2_C1$_{Cterm}$ to Txo2_C1$_{Nterm}$. The first one (α-crossover) is within the linker between β8 and β9 strands of Txo2_C1$_{Cterm}$, contributing a short helix (3_{10}-helix) to Txo2_C1$_{Nterm}$ (Fig. 2B and SI Fig. 2). The second one (β-crossover) is within the linker between β10 and β12 strands of Txo2_C1$_{Cterm,}$ donating typically two strands (such as β11-strand, Fig. 2A) onto one side of the central β-sheet of Txo2_C1$_{Nterm}$ (Drake et al., 2016; Bloudoff et al., 2013; Samel et al., 2007; Reimer et al., 2016). The edge strand of the two strands could be in the form of a coil in some structures (Keating et al., 2002; Haslinger et al., 2015). In Txo2_C1-A1, the edge strand (or coil) can't be modelled due to weak electron density. The β-crossover that bridges the two subdomains has been proposed as a possible "latch" that

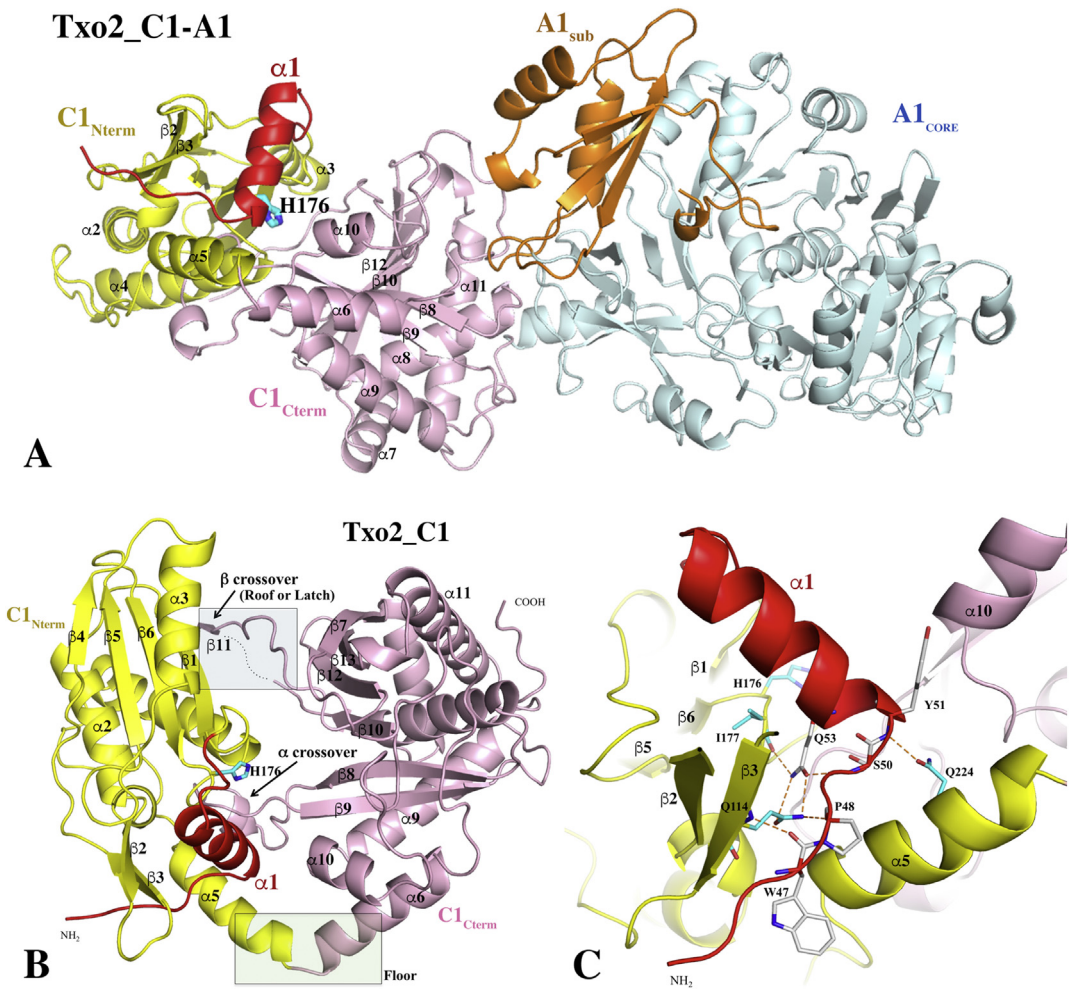

Fig. 2. Structure of Txo2_C1-A1 bidomain construct. (A) A ribbon diagram of Txo2_C1-A1 structure. The $C1_{Nterm}$ subdomain is colored in yellow, except for the α1 helix that is colored in red to highlight a different orientation. The $C1_{Cterm}$ subdomain is colored in pink. The $A1_{core}$ and $A1_{sub}$ domains are colored in light cyan and orange, respectively. The catalytic residue of C1 domain H176 is drawn in stick format to highlight its position. (B) The Txo2_C1 domain is re-oriented to show the tunnel between its two subdomains. The connection region between two subdomains is called the floor of the tunnel. Alongside the first crossover (α-crossover) from $C1_{Cterm}$ to $C1_{Nterm}$, the second crossover (β-crossover) is called the roof of the tunnel. The roof has been proposed to function as a "latch". The opening of the "latch" provides access to the catalytic site of the C domain. (C) The hydrogen bond interactions involved in the N-terminal region up to the α1 helix. For clarity, the hydrogen bond between S50 and Y271 are not shown. The hydrogen bonds displayed in figures or mentioned above have their bond distances less than 3.5 Å. All the residues involved in these interactions are conserved in Txo1_C2 domain. Figs. 2–6 and SI Figs. 3–8 are prepared with the program PyMOL (http://www.PyMOL.org).

forms the "roof" of the active site (Samel et al., 2007) (Fig. 2A). The opening and closing of this "latch" during catalysis of the C domain could regulate access of the substrates to the active sites residing in the middle of the tunnel.

The two C subdomains are hold together by α- and β-crossovers that also help to limit their relative movement. It is noticed that the interaction between the N-terminal α1 helix (from S50 to M62, SI Fig. 2) and the rest of the C domain is predominantly hydrophobic with an interface area of about 840 $Å^2$ (Krissinel & Henrick, 2007). The interface residues include Y51, A52, Q53, L56, W57 and L69 from the α1 helix, in which only Q53 contributes a few hydrogen bonds at the beginning of the α1 helix (Fig. 2C). This structural feature will be elaborated on later.

Like a typical A domain (Conti et al., 1997), Txo2_A1 is comprised of a large N-terminal subdomain (or A_{core} subdomain) and a small C-terminal subdomain (or A_{sub} subdomain), Fig. 2A. A_{core} can be further divided into three parts, characterized by two central β-sheets (A and B) and a distorted β-barrel (C) (Conti et al., 1997). In the Txo2_C1-A1 bidomain construct, the C and A domains are closely associated, with an interface area of about 553 $Å^2$ (Krissinel & Henrick, 2007) (SI Table I) without considering the contribution from the linker between the two domains (Q476 to L484). $Txo2\_C1_{Nterm}$ makes no contact to Txo2_A1. It is the $Txo2\_C1_{Cterm}$ that uses its α6_β7, α8_β8 and β9_α9 loops as well as its last helix, α11 (SI Fig. 1) for the C1-A1 contact. On the A domain side, about 53% (291 $Å^2$) of the bidomain interface is contributed by the $Txo2\_A1_{core}$ domain, Fig. 2A. The contribution of $Txo2\_A1_{core}$ primarily comes from the β-barrel C, including the C2_C3 loop, and C5_C6 loop and its proximities. The relative B-factor ratios of these subdomains ($C1_{Nterm}$: $C1_{Cterm}$: $A1_{core}$: $A1_{sub}$ = 1.83 : 1.00: 1.58 : 2.27) seemingly indicate higher mobility of $C1_{Nterm}$ and $A1_{sub}$ in respect to $C1_{Cterm}$ and $A1_{core}$.

Multiple conformations have been proposed for the reaction cycle catalyzed by A domain (Yonus et al., 2008; Strieker et al., 2010). The Txo2_A1 adopts a closed conformation (potentially adenylate-forming conformation) (Fig. 2A). Other structural models of representatives of an closed state include the structure of the N-terminal phenylalanine-activating A domain (PheA) of gramicidin S NRPSs (Conti et al., 1997) and the crystal structures of the stand-alone aryl acid activating domain of NRPS DhbE that initiates the bacillibactin synthesis (May et al., 2002). A superposition of Txo2_A1 and DhbE (PDB ID: 1MDF) (Krissinel & Henrick, 2004) yields a RMSD value of 2.5 Å with 444 residues from each A domain (~550 a.a.) aligned in spite of their low

sequence identity (20.5%) and different substrate specificities. The enterobactin synthetase EntF A domain is another structurally available serine specific A domain (Drake et al., 2016). The EntF A domain (PDB ID: 5T3D) is in a thioester-forming conformation with a rotation of its A_{sub} about 140° away from its position in an adenylate-forming conformation. When the A_{core} of the EntF A is aligned with Txo2_$A1_{core}$, the resulting RMSD value is 1.36 Å with 375 residues from each A_{core} (~420 a.a.) aligned, and a sequence identity of 47.7%.

3.2. Open conformation of C domain in Txo1_C2-$A3_{core}$ structure

Multiple constructs were designed to obtain the crystal structure of the serine module of Txo1, including Txo1_C2, _A3 and _PCP3 subunits. The construct that produced the best crystals had small A_{sub} subdomain deleted, which enabled us to determine its structure with SAD (Table 1). Subsequently, the structure of Txo1_C2-$A3_{core}$ was used as a template for the determination of other crystal structures, including Txo2_C1-A1 presented above.

In the crystal of Txo1_C2-$A3_{core}$, there is one monomer in the asymmetric unit (Fig. 3A). Interestingly, the monomer and one of its symmetry-related monomers swap their N-terminal α1 helices of Txo1_C2 domain (Fig. 3B), forming a dimer-like assembly. It is known that NRPSs are monomers in solution. The structure reveals the mobility of the α1 helix relative to the rest of the C domain. Txo1_C2 and Txo2_C1 both catalyze the addition of L-Ser to a growing peptide, but at different stages. Their acceptor substrate (L-Ser) is the same, while their donor substrates are different. The donor substrate of Txo2_C1 is four residues longer than that of Txo1_C2. The two C domains have 80.1% sequence identity (SI Fig. 2). When the divergent sequences are mapped onto the molecular surface of one C domain, the acceptor PCP domain docking site (Drake et al., 2016; Kreitler et al., 2019) is obviously more conserved than the docking site of donor PCP domain (Samel et al., 2007; Tarry et al., 2017) (SI Fig. 3). The two C domains can be superimposed with an RMSD value of 1.25 Å (Krissinel & Henrick, 2004) without apparent conformational changes between each other, excepting the orientation of the N-terminal α1 helix (Fig. 3C). However, if only $C2_{Cterm}$ is used for an alignment, a small rotation (~5°) of $C2_{Nterm}$ in relation to $C2_{Cterm}$ is observed between Txo1_C2 and Txo2_C1 (SI Fig. 4). $C2_{Nterm}$ also has a larger average B-factor in comparison to $C2_{Cterm}$ and $A3_{core}$ subdomains with ratios of $C2_{Nterm}$: $C2_{Cterm}$: $A3_{core}$ = 3.84 : 1.99: 1.00.

The α1 helix in the Txo1_C2 structure swings out (Fig. 3A–C), exposing the active site inside the tunnel of the C domain (SI Fig. 5). In the following discussions, we will refer the Txo1_C2 conformation as an open conformation of the C domain. For comparison, we will refer the conformation of the C1 domain in the Txo2_C1 structure described earlier as a closed conformation of C domain.

The unexpected α1 helix out-movement exposes the mostly hydrophobic interface between the helix and the rest of the C domain. Additionally, part of the loop following the α1 helix in Txo2_C1 (closed conformation) becomes helical; this consequently elongates the α1 helix

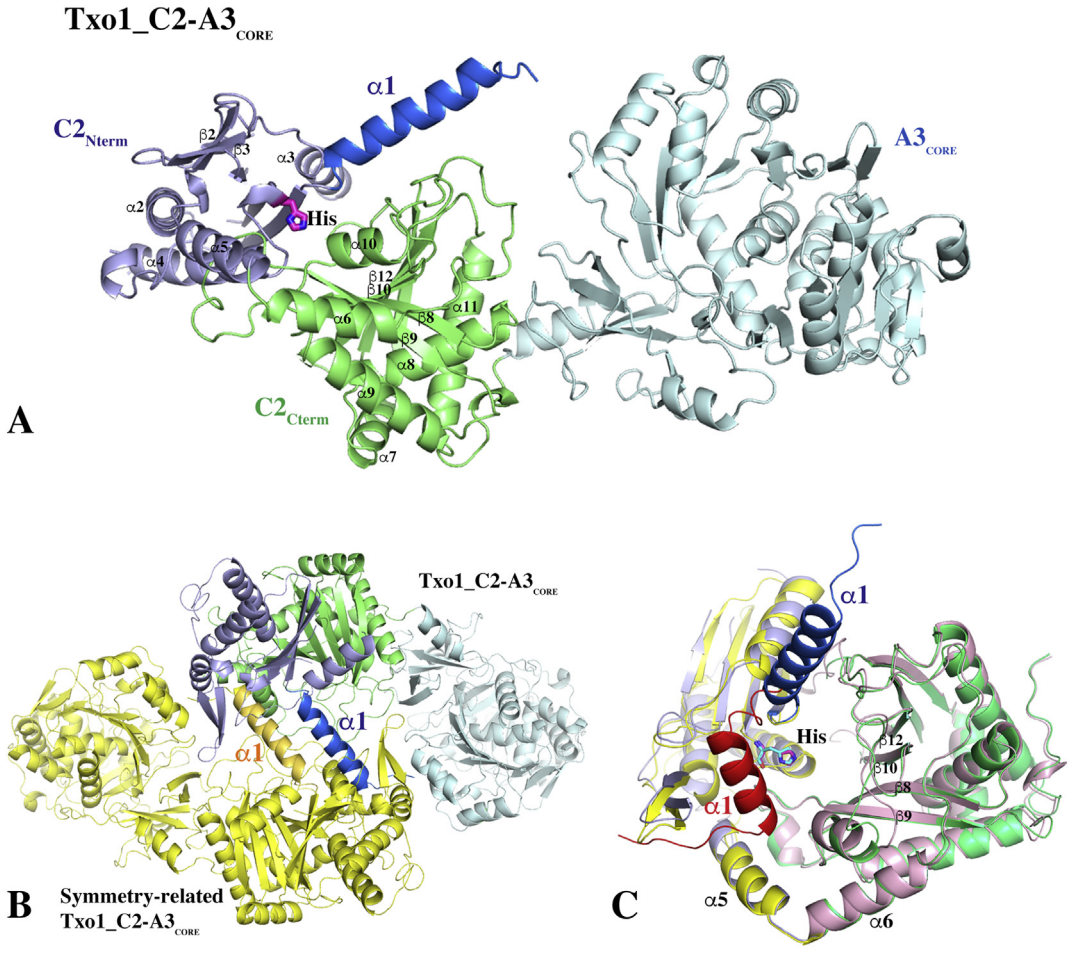

Fig. 3. Structure of Txo1_C2-$A3_{core}$ bidomain construct. (A) A ribbon diagram of Txo1_C2-$A3_{core}$ structure. The $C2_{Nterm}$ subdomain is colored in light blue except the α1 helix that is colored in blue for highlighting. The $C2_{Cterm}$ subdomain is colored in green. The $A3_{core}$ subdomains are colored in cyan. The catalytic residue of C2 domain H2268 is shown in stick format. (B) The swapping of α1 helix of the Txo1_C2-$A3_{core}$ with a symmetry-related molecule Txo1_C2-$A3_{core}$, which is colored in yellow with its α1 helix shown in dark yellow. The two monomers form a dimer-like assembly in crystal. (C) The structural alignment of Txo2_C1 and Txo1_C2 for illustration of the exchange of the α1 helix from its position in Txo2_C1 to its position in Txo1_C2.

from 3.3 turns to 5 turns in Txo1_C2 (open conformation) (see Fig. 3 and SI Fig. 2). This open conformation is stabilized by the swapping of the α1 helix between two symmetry-related monomers. This molecular packing implies that the opening of the C domain is likely transient, representing one of many steps in the cycle of peptide bond-formation activity of the C domain.

As mentioned earlier, substrates carried by donor PCP and acceptor PCP enter C domain from two sides of the domain for peptide elongation (Fig. 1A). The functional docking of acceptor PCP to C domain has been well studied (Drake et al., 2016; Kreitler et al., 2019), while the active interaction of donor PCP to C domain needs to be further investigated (Samel et al., 2007; Tarry et al., 2017). To examine the relevance of the α1 helix motion to substrate entrance/exit of C domain, a docking mode of acceptor PCP on Txo1_C2 domain is modelled from the four-domain (C-A-PCP-TE) structure of NRPS AB3403 (Drake et al., 2016), Fig. 4. In the AB3403 structure, the acceptor PCP domain interacts with C domain and inserts a phosphopantetheine cofactor attached to its pantetheinylation site into the active site of the C domain (Fig. 4A). One side of the tunnel for the co-factor and its attached substrate is formed by α1 helix, especially at the entrance of the tunnel on this side, Fig. 4D. The α1 helix apparently limits the size of tunnel, and thus limits the size of the molecules able to pass in and out of the active site. However, when α1 helix opens up, as seen in Txo1_C2 (Fig. 4E), access to the active site for substrate before condensation reaction or the exit from active site for elongated peptide after the reaction become much less limited.

3.3. A_{core} domain and AMP-binding in the absence of Mg^{2+}

Txo1_A3 and Txo2_A1 are nearly identical in sequence (99% identity, SI Fig. 1), they use the same substrate (L-Ser), and their core structures (Txo1_$A3_{core}$ and Txo2_$A1_{core}$) can be superimposed very well with an RMSD value of 0.65 Å. The absence of Txo1_$A3_{sub}$ in the Txo1_C2-A3 construct (Fig. 3A) apparently does not have any substantial impact on the conformation of Txo1_$A3_{core}$. However, Txo1_$A3_{core}$ moves closer to Txo1_C2 with a rotation of about 17° (SI Fig. 6). The interface area between Txo1_$C2_{Cterm}$ and Txo1 $A3_{core}$ increases to 371 $Å^2$ compared to the interface area (291 $Å^2$) between Txo2_$C1_{Cterm}$ and Txo1 $A1_{core}$ (SI Table I). It is not clear if the conformation change between C_{Cterm} and A_{core} in these C-A bidomain constructs is related to the presence/absence of A_{sub} domain. Though the C-A interdomain linker limits the relative motion of the two subdomains, their interacting surface area (excluding linker contribution) seems to be quite dynamic (SI Table I) (Bloudoff & Schmeing, 2017).

To examine the mode of substrate binding in an A domain lacking the A_{sub} subdomain, particularly in the absence of the strictly conserved A_{sub} lysine residue (K3091 in Txo1_A3, SI Fig. 1), we further co-crystallized Txo1_C2_$A3_{core}$ with L-Ser and AMP. The invariant lysine is believed to be important in stabilizing and activating substrate (or intermediate) (Conti et al., 1997; Gulick et al., 2003). Although no serine was observed in the substrate-binding site, the electron density for AMP is well-defined (Fig. 5A and B) indicating that A_{sub} (including K3091) may play role in positioning the substrate serine.

The AMP-binding mode in Txo1_C2_$A3_{core}$ is quite similar to what was reported for other AMP-binding A domains, particularly for the adenosine moiety (Fig. 5A and B) (Conti et al., 1997; May et al., 2002). Furthermore, the absence of A_{sub} does not affect AMP binding mode within A_{core}. The adenine ring of AMP is bound inside a largely hydrophobic pocket, surrounded by the side chains of two tyrosines (Y2893 and Y2997), and a loop (in red) containing a sequence motif of S2868GER (Fig. 5B). At the bottom of the pocket, the nitrogen atom N6 of adenine forms a hydrogen bond with the side chain of conserved asparagine N2891 that is part of strand B7. A water bridge is also found between the N1 of adenine and asparagine N2891. The two interactions are well conserved in all reported A domain complexes with AMP/ATP (Conti et al., 1997; May et al., 2002; Yonus et al., 2008; Mitchell et al., 2012). There is a hydrogen bond between the N6 of adenine and a main

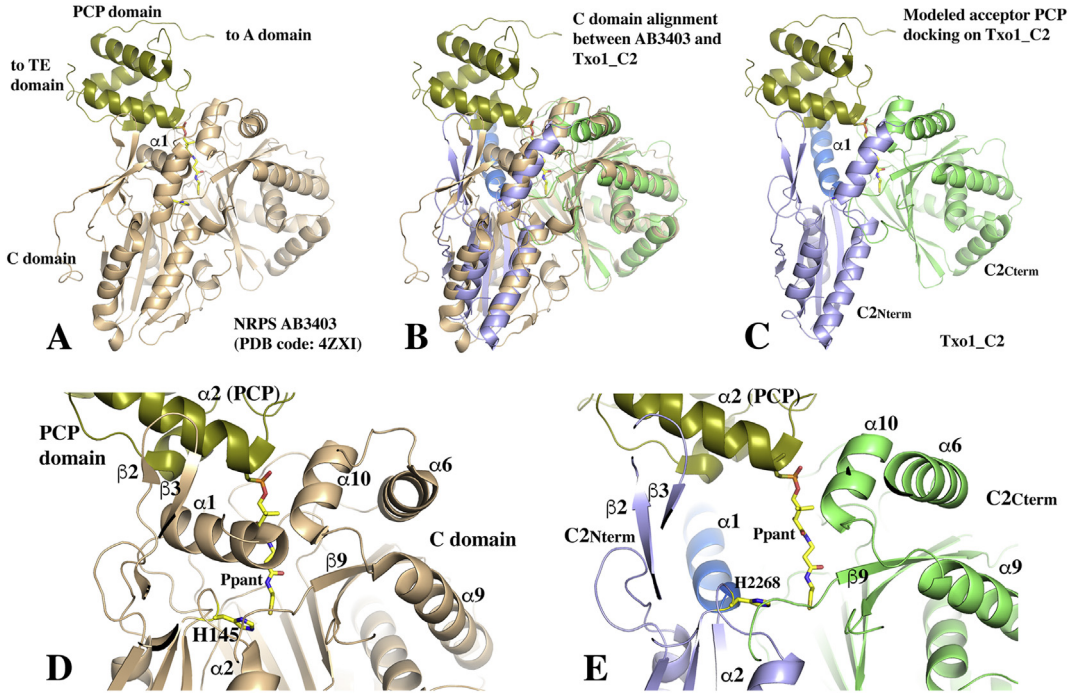

Fig. 4. Modeling of acceptor PCP domain docking on Txo1_C2. (A) The docking of acceptor PCP domain on C domain from the crystal structure of NRPS AB3403 (PDB ID: 4ZXI) (Drake et al., 2016). (B) A C domain structural alignment between AB3403 and Txo1_C2. The alignment results in an RMSD value of 2.25 Å with 338 out of about 410 residues from each one aligned and a sequence identity of 24.9%. (C) A docking model of acceptor PCP domain on Txo1_C2 when AB3403 C domain is replaced by Txo1_C2. (D) Zoom-in view of the acceptor PCP docking on C domain in AB3403. The phosphopantetheine (Ppant) attached to PCP domain and the catalytic histidine are drawn in stick format. (E) Zoom-in view of the acceptor PCP docking on Txo1_C2 domain. The swiping out of α1 helix apparently opens up the tunnel to the active site of C domain.

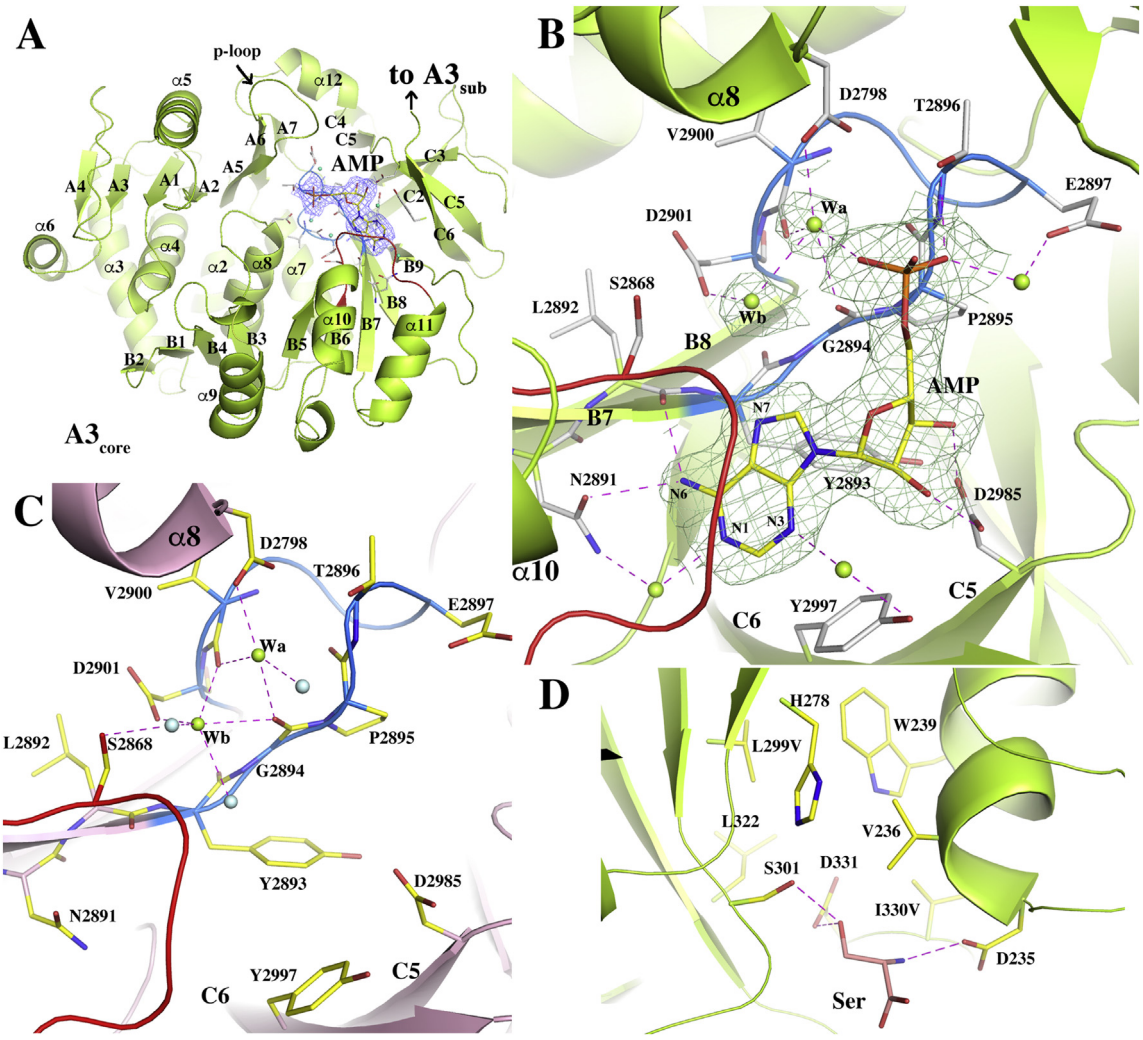

Fig. 5. AMP-binding in Txo1_C2-A3$_{core}$ subdomain in the absence of Mg^{2+}. (A) A ribbon diagram of Txo1_A3$_{core}$ in complex with AMP. A $2F_o$-F_c difference electron density map for AMP is drawn in blue mesh at 1σ contour level to highlight its position between the β-sheet B and the distorted β-barrel C. The labeling of the secondary structures of the Txo1_A3$_{core}$ largely follows that described in PheA structure (Conti et al., 1997). Two loops, the red one between the strand B6 and the helix α11 and the cyan one between the strands B7 and B8, are highlighted. The p-loop (Saraste et al., 1990) between the strands A5 and A6 is indicated by an arrow. (B) AMP-binding site in Txo1_C2-A3$_{core}$ subdomain. Key residues and water molecules involved in AMP-binding are represented by stick and sphere formats, respectively. For clarity, the $2F_o$-F_c difference electron density map at countour level of 1σ is shown for AMP and two water molecules. (C) The AMP-binding site in Txo1_C2-A3$_{core}$ in the absence of AMP, showing solvent structure. (D) A modeling of serine binding in A domain based on an A domain structural alignment with PheA (PDB ID: 1AMU) (Conti et al., 1997). All hydrogen bonds displayed in figures have their bond distances less than 3.5 Å.

chain carbonyl oxygen (L2892). There is also a water bridge between the N3 of adenine and the hydroxyl group of Y2997 (Fig. 5B). The AMP ribose ring forms two hydrogen bonds with the side chain of the highly conserved aspartate (D2985) that is part of strand C5 (Fig. 5B). These two hydrogen bonds are also well-conserved in the A domain complexes with AMP/ATP (Conti et al., 1997; May et al., 2002; Yonus et al., 2008; Mitchell et al., 2012). The structure of Txo1_C2-A3$_{core}$/AMP complex reported here indicates that the positioning of AMP (or ATP) doesn't require the presence of A_{sub}.

The α-phosphate group of AMP forms a hydrogen bond with conserved threonine T2896, and is water-bridged to the invariant glutamate (E2897) (Fig. 5B) (Marahiel et al., 1997). The two residues are in the loop (in cyan) between the strands B7 and B8. The B7_B8 loop also harbors a well-conserved GP sequence motif (G2894P, see SI Fig. 1) (Marahiel et al., 1997). The presence of the proline residue forces the carbonyl group of the glycine to point up with a rotation of about 60° (Fig. 5B and C). Consequently, this carbonyl group, with the carbonyl group from a residue on the other side of the loop (V2900), along with the carboxyl group of an aspartate (D2798) create an environment for binding a positively charged atom such as a cation like Mg^{2+} or an amide group from a substrate as showed below.

Since the substrate serine was not observed in structure though the amino acid was used in co-crystallization, a model was created based on an alignment of Txo1_A3$_{core}$ to a known substrate-binding A domain (PheA) (Conti et al., 1997) to evaluate a possible binding mode of the serine in Txo1_A3 and Txo2_A1, which are both serine-specific A domains. In the model shown in Fig. 5D, the selectivity-conferring residues (Stachelhaus code) (Stachelhaus et al., 1999) seem to define a pocket larger than the size of the small amino acid substrate. The side chain of the serine could potentially form hydrogen bonds with S301/S2869 (Stachelhaus code numbering/Txo1) and/or D331/D2901 with a bond length of about 2.6 Å. As discussed above, the amide group of the amino acid interacts with highly conserved D235/D2798 (Stachelhaus code numbering/Txo1) with a similar bond distance. The carbonyl group of serine could interact with the α-phosphate group of a bound ATP/AMP

(not shown in the figure).

3.4. A_{core} domain and its AMP-binding in the presence of Mg^{2+}

The position of the Mg^{2+} required for adenylation reactions (Airas, 2007; Schmelz & Naismith, 2009), has been ambiguous since the first reported A domain structure (Conti et al., 1997). Txo1_C2-A3$_{core}$ crystals obtained from Mg^{2+}-containing crystallization solutions contained bound Mg^{2+}ion in the structure. The Mg^{2+} ion displaced the water molecules discussed earlier (W_a and W_b in Fig. 5C) (Fig. 6A). In addition to two carbonyl groups from G2894 and V2900, three water molecules contribute to magnesium coordination (W_1, W_2 and W_3), while the sixth position of the Mg^{2+} octahedral coordination sphere is occupied by an oxygen atom from the sulfate group of a MES molecule, a component of the crystallization buffer (Fig. 6A). The similarity of a sulfate group to a phosphate group implies that the Mg^{2+} could potentially interact with a nucleotide phosphate group.

When the structure of Txo1_C2-A3$_{core}$/AMP complex crystal grown from a Mg^{2+}-containing buffer was determined, the α-phosphate group of AMP was indeed found interacting with Mg^{2+} (Fig. 6B). Compared to the AMP-binding in the absence of Mg^{2+} (Fig. 5B), a bending of the α-phosphate group towards to Mg^{2+}-binding site was observed without significant impact to the interaction pattern of the adenine ring with its surroundings (SI Fig. 7). Only the water that bridges the N3 of adenine and the hydroxyl group of Y2997 in the structure of Txo1_C2-A3$_{core}$/AMP (Fig. 5B) moved more than 0.4 Å away from the N3 atom. The well-defined Mg^{2+}-binding site observed in the Txo1_C2-A3$_{core}$/AMP/Mg^{2+} structure and interaction with AMP (a product of acetylation), we believe, may represent part of an intermediate state of the ATP- and Mg^{2+}-dependent catalysis of A domain.

4. Discussions

A conformational change of a C domain, involving an opening of a "latch" or a "lid" above the tunnel between two subdomains, had been proposed years ago (Samel et al., 2007). The assumption was based on the requirement that both donor and acceptor substrates need access to the active site for the reaction to proceed (Fig. 1B). As described earlier, there are two crossovers between C_{Cterm} and C_{Nterm} subdomainsan α-crossover and a β-crossover (Fig. 2B). The β-crossover that covers the top of tunnel was initially proposed as a possible "latch" for the opening of the tunnel (Bloudoff et al., 2013; Samel et al., 2007). Such β-crossovers resemble a lid covering an active site, as observed in the structures of NRPS peptide-cyclizing thioesterase domains (Bruner et al., 2002; Samel et al., 2006) as well as in common lipases (Holmquist, 2000). However, the opening of the "latch" in NRPS C domains has never been observed and the β-crossover is quite divergent in both sequence and structure (SI Fig. 2). In some structures of C domain, the edge strand of the β-crossover becomes part of a loop (Keating et al., 2002; Tanovic et al., 2008; Haslinger et al., 2015), indicating its high mobility.

The α-crossover together with β-crossover apparently help hold the two subdomains together in a specific conformation (Fig. 2B). Whether the latch, the β-crossover, could be opened so that substrates can access the active site remains to be determined. From all known structurally characterized C domains, no significant orientation change between two subdomains has ever been observed for a given C domain. Only limited orientation differences between two subdomains in different C domains were described when they are superimposed with an alignment of their C_{Cterm} subdomains (Bloudoff et al., 2013).

Though a partially degenerated α1 helix was found in NRPS EntF C structures (Miller et al., 2016), the open conformation of C domain in the Txo1_C2-A3$_{core}$, if functionally validated, may represent the first observation of a conformational change of any structurally known C domains. It is possible that the open conformation of the α1 helix represents a state within a C domain catalytic cycle. The interface between α1 helix and the rest of condensation domain is predominantly hydrophobic, as described earlier. This provides a structural basis for the mobility of the helix and its potential functional regulation, as observed in other proteins. We have noticed the C-terminal α7 helix of integrin I-domain exhibits a similar predominantly hydrophobic interaction with the rest of I-domain. The α7 helix moves down and up along a side of I-domain, mediating the opening and closing of the metal ion-dependent adhesion site (MIDAS)

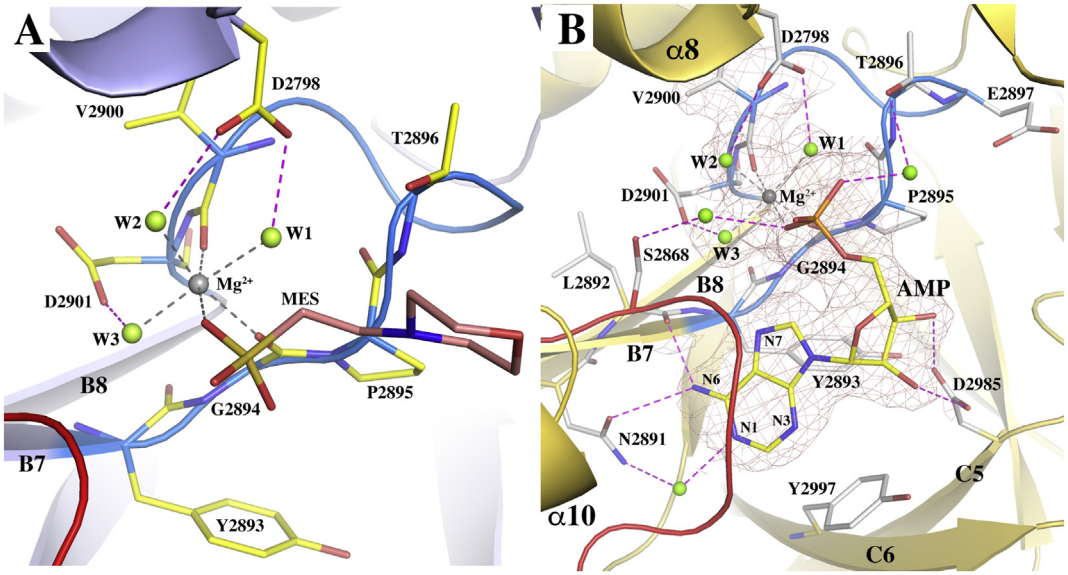

Fig. 6. Mg^{2+}-binding site and its interaction with AMP. (A) The Mg^{2+}-binding site associated with the loop between B7 and B8 strands. For clarity, electron density map is not shown. Key residues in the binding site are drawn in stick format. An anionic 2-morpholinoethanesulfonic acid (MES) molecule from crystallization buffer is also drawn in stick format. The morpholine ring of MES is partially disordered. The three water molecules (W1, W2 and W3) coordinated to Mg^{2+} are drawn as green spheres. The six coordinate bonds to Mg^{2+} are drawn as grey dashed lines with their bond lengths ranging from 2.02 Å to 2.21 Å. No bond geometry restraint was applied in structural refinement. The three hydrogen bonds with three water molecules are drawn in magenta dashed line. (B) The interactions between Mg^{2+} and AMP. A $2F_o$-F_c difference electron density map at 1σ contour level is drawn as the pink mesh for Mg^{2+}, AMP and three water molecules to demonstrate the quality of the map. All hydrogen bonds associated with AMP binding are drawn in magenta dashed line. The bond length between the six coordinating oxygen atoms and Mg^{2+} ion range from 1.99 Å to 2.19 Å without geometry restraints used in structural refinement.

and dramatically affecting the binding affinity of MIDAS to ligand (Luo et al., 2007). Remarkably, the α7 helix of I-domain has also been observed to swing out in an α_L integrin I domain complex with ICAM-5 (Zhang et al., 2008). The swung out α7 helix inserts into the α7 helix site of a neighboring symmetry-related I-domain in crystal packing, which is analogous to the α1 helix swapping between two symmetry-related Txo1_C2-A3$_{core}$ molecules in the crystal packing arrangement. In both cases, the swinging out of the domains' terminal helices demonstrates the high mobility that would allow them to serve a regulatory function. Therefore, we believe that the open conformation observed in Txo1_C2-A3$_{core}$ structure may represent a long-awaited open conformation of C domain, and that the outward movement of its N-terminal helix represents the transition from a closed to an open conformation. It is worth mentioning that a regulatory helix in amidase AmiB is also believed to block/open access to the active site of the enzyme by switching its positions (Yang et al., 2012). Interestingly, the interaction between this helix and the active site is also predominantly hydrophobic. Another example of molecular switching that involves displacement of a helix relative to its core domain includes light-activated kinase phototropin (Harper et al., 2003).

In NRPSs, A domain catalyzes the adenylation of an amino acid substrate through the formation of a phosphodiester bond between a hydroxyl group of the amino acid and an AMP derived from an ATP. A metal ion, usually Mg^{2+} is required for the activity of adenylate-forming enzymes (Airas, 2007; Schmelz & Naismith, 2009). Aside from its role in neutralizing the charge of ATP and the leaving group pyrophosphate, the metal ion is also expected to stabilize the negatively charged pentavalent phosphorus, a critical intermediate in the catalytic cycle. However, the number of metal ions necessary for the reaction and their possible binding-sites in an A domain had remained unanswered even after reports of several other A domain structures (Conti et al., 1997; Drake et al., 2016; Kaljunen et al., 2015).

The observation of a well-defined Mg^{2+}-binding site in the Txo1_A3$_{core}$ subdomain provides valuable insight. First, the binding site is associated with the loop between the B7 and B8 strands. A part of the loop itself was regarded as one of the functional coding regions of the A domain (Stachelhaus et al., 1999). Therefore, we believe this region helps define the Mg^{2+}-binding site of A domain. Second, the Mg^{2+} from the binding site can interact with the α-phosphate group of a bound AMP or ATP. It can also potentially interact with the β-phosphate group of ATP. Therefore, it is possible that the Mg^{2+} can play the important role in neutralizing the negative charge of ATP and the leaving pyrophosphate. Moreover, the positioned Mg^{2+} would be able to stabilize the intermediate pentavalent phosphorus at the center of adenylation reaction.

Interestingly, when we superimpose the Mg^{2+}-binding Txo1_A3$_{core}$ with some substrate-binding A domains, the Mg^{2+} site essentially overlaps with the amide group of phenylalanine (PDB ID:1AMU) (Conti et al., 1997), valine (PDB ID:3VNS), or glycine (PDB ID:4ZXI) (Drake et al., 2016), presumably hydrolyzed from phenylalanyl- or valyl- or glycyl-adenylate, respectively (SI Fig. 8). The amide of the valine moiety from inhibitor Val-AVS (valine-adenosine vinylsulfonamide) is also positioned at the Mg^{2+} site (PDB ID:4DG9) (Mitchell et al., 2012) (SI Fig. 8). It is also true for the α-amino acid substrate adenylation enzyme (PDB ID:3WV5) (Miyanaga et al., 2014). We have noticed that Mg^{2+} was assigned at the same site in the structures of a bi-specific A domain from anabaenopeptin synthetase in complex with ATP analog ANP (phosphoaminophosphonic acid-adenylate ester) (PDB IDs: 4D4G and 4D4I) (Kaljunen et al., 2015). However, the density for the metal ion and its coordination geometry are far from optimal in either case.

Surveying the corresponding Mg^{2+}-binding site in known A domain structures (Conti et al., 1997; Drake et al., 2016; Yonus et al., 2008; Mitchell et al., 2012; Tanovic et al., 2008; Reimer et al., 2016; Kaljunen et al., 2015) revealed several properties. In apo forms of the A domain, the metal-binding site is unoccupied such as in the apo Txo1_C2_A3$_{core}$ structure (Table 1), even when water molecules are present nearby. When a substrate or an inhibitor is present in structure, the positively charged amide group of the substrate or the inhibitor will occupy this site (SI Fig. 6). However, in the absence of substrate or inhibitor, and when Mg^{2+} is present, the metal ion will move into the site as in the Txo1_C2_A3$_{core}$/Mg^{2+} structures (Table 1). The mutual exclusion of Mg^{2+} and the amide from a substrate at the same site in a structure seems to indicate that these two structures, substrate-occupied and Mg^{2+}-occupied, may correspond to two separate states of the A domain adenylation reaction.

In summary, we believe the Mg^{2+}-binding site identified in Txo1_C2_A3$_{core}$ structures provides insights into the catalytic mechanism of the adenylation reaction carried by A domain. However, to answer the question why this Mg^{2+}-binding mode was never observed in other A domain structures will require further investigation.

Author contributions

KT, MZ, RPJ, RW, RAH and DB performed experiments and structure determination, MZ and RW purified, characterized and crystallized proteins for biochemical and biophysical studies, RPJ cloned genes and expressed proteins, GB performed bioinformatic analysis and designed constructs, AJ, KT, DB designed experiments, analyzed data, and together wrote manuscript.

Declaration of Competing Interest

Authors declare no conflict of interest.

Acknowledgments

This work was supported by Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH), Department of Health and Human Services, under Contract Nos. HHSN272201200026C and HHSN272201700060C (to A. J). We thank members of the Structural Biology Center at Argonne National Laboratory for their help in conducting X-ray diffraction data collection. BUILDing SCHOLARS Summer Research Program is supported by the National Institute of General Medical Sciences of the National Institutes of Health under linked Award Numbers RL5GM118969, TL4GM118971, and UL1GM118970. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The Structural Biology Center beamlines are supported by U.S. Department of Energy, Office of Biological and Environmental Research, under contract DE-AC02-06CH11357. We are grateful to Chase Akins for his proofreading.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crstbi.2020.01.002.

References

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., et al., 2012. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. 68, 352–367.

Agrawal, S., Acharya, D., Adholeya, A., Barrow, C.J., Deshmukh, S.K., 2017. Nonribosomal peptides from marine microbes and their antimicrobial and anticancer potential. Front. Pharmacol 8, 828.

Airas, R.K., 2007. Magnesium dependence of the measured equilibrium constants of aminoacyl-tRNA synthetases. Biophys. Chem. 131, 29–35.

Aslanidis, C., de Jong, P.J., 1990. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. 18, 6069–6074.

Bloudoff, K., Alonzo, D.A., Schmeing, T.M., 2016. Chemical probes allow structural insight into the condensation reaction of nonribosomal peptide synthetases. Cell Chem Biol 23, 331–339.

Bloudoff, K., Rodionov, D., 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.

Bloudoff, K., Schmeing, T.M., 2017. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. Biochim. Biophys. Acta Protein Proteonomics 1865, 1587–1604.
Blommel, P.G., Fox, B.G., 2007. A combined approach to improving large-scale production of tobacco etch virus protease. Protein Expr. Purif. 55, 53–68.
Bruner, S.D., Weber, T., Kohli, R.M., Schwarzer, D., Marahiel, M.A., Walsh, C.T., et al., 2002. Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. Structure 10, 301–310.
Challis, G.L., Naismith, J.H., 2004. Structural aspects of non-ribosomal peptide biosynthesis. Curr. Opin. Struct. Biol. 14, 748–756.
Chen, W.H., Li, K., Guntaka, N.S., Bruner, S.D., 2016. Interdomain and intermodule organization in epimerization domain containing nonribosomal peptide synthetases. ACS Chem. Biol. 11, 2293–2303.
Conti, E., Stachelhaus, T., Marahiel, M.A., Brick, P., 1997. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. EMBO J. 16, 4174–4183.
Dehling, E., Volkmann, G., Matern, J.C., Dorner, W., Alfermann, J., Diecker, J., et al., 2016. Mapping of the communication-mediating interface in nonribosomal peptide synthetases using a genetically encoded photocrosslinker supports an upside-down helix-hand motif. J. Mol. Biol. 428, 4345–4360.
Dowling, D.P., Kung, Y., Croft, A.K., Taghizadeh, K., Kelly, W.L., Walsh, C.T., et al., 2016. Structural elements of an NRPS cyclization domain and its intermodule docking domain. Proc. Natl. Acad. Sci. U. S. A. 113, 12432–12437.
Drake, E.J., Miller, B.R., Shi, C., Tarrasch, J.T., Sundlov, J.A., Allen, C.L., et al., 2016. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. Nature 529, 235–238.
Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132.
Eschenfeldt, W.H., Lucy, S., Millard, C.S., Joachimiak, A., Mark, I.D., 2009. A family of LIC vectors for high-throughput cloning and purification of proteins. Methods Mol. Biol. 498, 105–115.
Eschenfeldt, W.H., Maltseva, N., Stols, L., Donnelly, M.I., Gu, M., Nocek, B., et al., 2010. Cleavable C-terminal His-tag vectors for structure determination. J. Struct. Funct. Genom. 11, 31–39.
Fischbach, M.A., Walsh, C.T., 2006. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. Chem. Rev. 106, 3468–3496.
Gulick, A.M., 2016. Structural insight into the necessary conformational changes of modular nonribosomal peptide synthetases. Curr. Opin. Chem. Biol. 35, 89–96.
Gulick, A.M., 2017. Nonribosomal peptide synthetase biosynthetic clusters of ESKAPE pathogens. Nat. Prod. Rep. 34, 981–1009.
Gulick, A.M., Starai, V.J., Horswill, A.R., Homick, K.M., Escalante-Semerena, J.C., 2003. The 1.75 A crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A. Biochemistry 42, 2866–2873.
Girt, G.C., Mahindra, A., Al Jabri, Z.J.H., De Ste Croix, M., Oggioni, M.R., Jamieson, A.G., 2018. Lipopeptidomimetics derived from teixobactin have potent antibacterial activity against Staphylococcus aureus. Chem. Commun. (Camb). 54, 2767–2770.
Hahn, M., Stachelhaus, T., 2006. Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. Proc. Natl. Acad. Sci. U. S. A. 103, 275–280.
Harper, S.M., Neil, L.C., Gardner, K.H., 2003. Structural basis of a phototropin light switch. Science 301, 1541–1544.
Haslinger, K., Peschke, M., Brieke, C., Maximowitsch, E., Cryle, M.J., 2015. X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. Nature 521, 105–109.
Holmquist, M., 2000. Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. Curr. Protein Pept. Sci. 1, 209–235.
Homma, T., Nuxoll, A., Gandt, A.B., Ebner, P., Engels, I., Schneider, T., et al., 2016. Dual targeting of cell wall precursors by teixobactin leads to cell lysis. Antimicrob. Agents Chemother. 60, 6510–6517.
Kaljunen, H., Schiefelbein, S.H., Stummer, D., Kozak, S., Meijers, R., Christiansen, G., et al., 2015. Structural elucidation of the bispecificity of A domains as a basis for activating non-natural amino acids. Angew Chem. Int. Ed. Engl. 54, 8833–8836.
Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797.
Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D Biol. Crystallogr. 60, 2256–2268.
Keating, T.A., Marshall, C.G., Walsh, C.T., Keating, A.E., 2002. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. Nat. Struct. Biol. 9, 522–526.
Kreitler, D.F., Gemmell, E.M., Schaffer, J.E., Wencewicz, T.A., Gulick, A.M., 2019. The structural basis of N-acyl-alpha-amino-beta-lactone formation catalyzed by a nonribosomal peptide synthetase. Nat. Commun. 10, 3432.
Leslie, A.G., Moody, P.C., Shaw, W.V., 1988. Structure of chloramphenicol acetyltransferase at 1.75-A resolution. Proc. Natl. Acad. Sci. U. S. A. 85, 4133–4137.
Ling, L.L., Schneider, T., Peoples, A.J., Spoering, A.L., Engels, I., Conlon, B.P., et al., 2015. A new antibiotic kills pathogens without detectable resistance. Nature 517, 455–459.
Luo, B.H., Carman, C.V., Springer, T.A., 2007. Structural basis of integrin regulation and signaling. Annu. Rev. Immunol. 25, 619–647.
Mandalapu, D., Ji, X., Chen, J., Guo, C., Liu, W.Q., Ding, W., et al., 2018. Thioesterase-mediated synthesis of teixobactin analogues: mechanism and substrate specificity. J. Org. Chem. 83, 7271–7275.
Marahiel, M.A., Stachelhaus, T., Mootz, H.D., 1997. Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem. Rev. 97, 2651–2674.
May, J.J., Kessler, N., Marahiel, M.A., Stubbs, M.T., 2002. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. Proc. Natl. Acad. Sci. U. S. A. 99, 12120–12125.
Miller, B.R., Gulick, A.M., 2016. Structural biology of nonribosomal peptide synthetases. Methods Mol. Biol. 1401, 3–29.
Minor, W., Cymborowski, M., Otwinowski, Z., Chruszcz, M., 2006. HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. Acta Crystallogr. D Biol. Crystallogr. 62, 859–866.
Mitchell, C.A., Shi, C., Aldrich, C.C., Gulick, A.M., 2012. Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. Biochemistry 51, 3252–3263.
Mootz, H.D., Marahiel, M.A., 1997. The tyrocidine biosynthesis operon of Bacillus brevis: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. J. Bacteriol. 179, 6843–6850.
Miller, B.R., Drake, E.J., Shi, C., Aldrich, C.C., 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.
Miyanaga, A., Cieslak, J., Shinohara, Y., Kudo, F., Eguchi, T., 2014. The crystal structure of the adenylation enzyme VinN reveals a unique beta-amino acid recognition mechanism. J. Biol. Chem. 289, 31448–31457.
Payne, J.A., Schoppet, M., Hansen, M.H., Cryle, M.J., 2016. Diversity of nature's assembly lines - recent discoveries in non-ribosomal peptide synthesis. Mol. Biosyst. 13, 9–22.
Reimer, J.M., Aloise, M.N., Harrison, P.M., Schmeing, T.M., 2016. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. Nature 529, 239–242.
Rosenbaum, G., Alkire, R.W., Evans, G., Rotella, F.J., Lazarski, K., Zhang, R.G., et al., 2006. The Structural Biology Center 19ID undulator beamline: facility specifications and protein crystallographic results. J. Synchrotron Radiat. 13, 30–45.
Schmelz, S., Naismith, J.H., 2009. Adenylate-forming enzymes. Curr. Opin. Struct. Biol. 19, 666–671.
Stachelhaus, T., Mootz, H.D., Marahiel, M.A., 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem. Biol. 6, 493–505.
Strieker, M., Tanovic, A., Marahiel, M.A., 2010. Nonribosomal peptide synthetases: structures and dynamics. Curr. Opin. Struct. Biol. 20, 234–240.
Samel, S.A., Schoenafinger, G., Knappe, T.A., Marahiel, M.A., Essen, L.O., 2007. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. Structure 15, 781–792.
Stols, L., Millard, C.S., Dementieva, I., Donnelly, M.I., 2004. Production of selenomethionine-labeled proteins in two-liter plastic bottles for structure determination. J. Struct. Funct. Genom. 5, 95–102.
Schneider, T.R., Sheldrick, G.M., 2002. Substructure solution with SHELXD. Acta Crystallogr. D Biol. Crystallogr. 58, 1772–1779.
Samel, S.A., Wagner, B., Marahiel, M.A., Essen, L.O., 2006. The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide. J. Mol. Biol. 359, 876–889.
Saraste, M., Sibbald, P.R., Wittinghofer, A., 1990. The P-loop–a common motif in ATP- and GTP-binding proteins. Trends Biochem. Sci. 15, 430–434.
Tanovic, A., Samel, S.A., Essen, L.O., Marahiel, M.A., 2008. Crystal structure of the termination module of a nonribosomal peptide synthetase. Science 321, 659–663.
Tarry, M.J., Haque, A.S., Bui, K.H., Schmeing, T.M., 2017. X-ray crystallography and electron microscopy of cross- and multi-module nonribosomal peptide synthetase proteins reveal a flexible architecture. Structure 25, 783–793 e4.
Vagin, A., Teplyakov, A., 2010. Molecular replacement with MOLREP. Acta Crystallogr. D Biol. Crystallogr. 66, 22–25.
Walsh, C.T., 2015. A chemocentric view of the natural product inventory. Nat. Chem. Biol. 11, 620–624.
Weissman, K.J., 2015. The structural biology of biosynthetic megaenzymes. Nat. Chem. Biol. 11, 660–670.
Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., et al., 2018. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. 27, 293–315.
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., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S., 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr., Sect D: Biol. Crystallogr. 67, 235–242.
Yang, D.C., Tan, K., Joachimiak, A., Bernhardt, T.G., 2012. A conformational switch controls cell wall-remodelling enzymes required for bacterial cell division. Mol. Microbiol. 85, 768–781.
Yonus, H., Neumann, P., Zimmermann, S., May, J.J., Marahiel, M.A., 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.
Yoon, J.R., Laible, P.D., Gu, M., Scott, H.N., Collart, F.R., 2002. Express primer tool for high-throughput gene cloning and expression. Biotechniques 33, 1328–1333.
Zhang, H., Casasnovas, J.M., Jin, M., Liu, J.H., Gahmberg, C.G., Springer, T.A., et al., 2008. An unusual allosteric mobility of the C-terminal helix of a high-affinity alphaL integrin I domain variant bound to ICAM-5. Mol. Cell 31, 432–437.
Zhang, J., Liu, N., Cacho, R.A., Gong, Z., Liu, Z., Qin, W., et al., 2016. Structural basis of nonribosomal peptide macrocyclization in fungi. Nat. Chem. Biol. 12, 1001–1003.