

# Structure of the adenylation–peptidyl carrier protein didomain of the *Microcystis aeruginosa* microcystin synthetase McyG

Xiao-Feng Tan, Ya-Nan Dai,\* Kang Zhou, Yong-Liang Jiang, Yan-Min Ren, Yuxing Chen and Cong-Zhao Zhou\*

Received 11 September 2014  
 Accepted 27 January 2015

Hefei National Laboratory for Physical Sciences at the Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230027, People's Republic of China. \*Correspondence e-mail: daiyanan@mail.ustc.edu.cn, zcz@ustc.edu.cn

Edited by Q. Hao, University of Hong Kong

**Keywords:** microcystin; NRPS/PKS; McyG; adenylation–peptidyl carrier protein didomain; reaction cycle.

**PDB reference:** McyG A–PCP didomain complexed with L-Phe-AMP, 4r0m

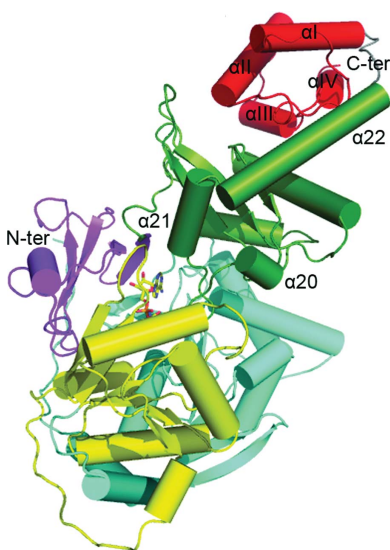
**Supporting information:** this article has supporting information at journals.iucr.org/d

Microcystins, which are the most common cause of hepatotoxicity associated with cyanobacterial water blooms, are assembled *in vivo* on a large multienzyme complex *via* a mixed nonribosomal peptide synthetase/polyketide synthetase (NRPS/PKS). The biosynthesis of microcystin in *Microcystis aeruginosa* PCC 7806 starts with the enzyme McyG, which contains an adenylation–peptidyl carrier protein (A–PCP) didomain for loading the starter unit to assemble the side chain of an Adda residue. However, the catalytic mechanism remains unclear. Here, the 2.45 Å resolution crystal structure of the McyG A–PCP didomain complexed with the catalytic intermediate L-phenylalanyl-adenylate (L-Phe-AMP) is reported. Each asymmetric unit contains two protein molecules, one of which consists of the A–PCP didomain and the other of which comprises only the A domain. Structural analyses suggest that Val227 is likely to be critical for the selection of hydrophobic substrates. Moreover, two distinct interfaces demonstrating variable crosstalk between the PCP domain and the A domain were observed. A catalytic cycle for the adenylation and peptide transfer of the A–PCP didomain is proposed.

## 1. Introduction

Cyanobacterial water blooms are of worldwide concern owing to their production of a wide range of hepatotoxins and neurotoxins. Of the known toxins, the hepatotoxic microcystins are produced by a diverse range of cyanobacteria, including species of the genera *Microcystis*, *Anabaena*, *Nostoc* and *Oscillatoria* (Rinehart *et al.*, 1994). It is known that acute poisoning by microcystins leads to death owing to hepatic haemorrhage (Jochimsen *et al.*, 1998; Pouria *et al.*, 1998), whereas the ingestion of sublethal amounts of microcystins increases the incidence rate of liver cancer (Nishiwaki-Matsushima *et al.*, 1992). Microcystins, which contain the novel aromatic  $\beta$ -amino-acid residue [(2*S*,3*S*,8*S*,9*S*)-3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-decadienoic acid; Adda], can inactivate serine/threonine protein phosphatase 1/2A (PP1/2A) in eukaryotes (Honkanen *et al.*, 1990) by specifically identifying the hydrophobic groove of PP1/2A (Goldberg *et al.*, 1995).

Previous biochemical and genetic studies have shown that microcystins are assembled by a mixed polyketide synthetase/nonribosomal peptide synthetase (PKS/NRPS; Moore *et al.*, 1991; Dittmann *et al.*, 1997) encoded by the microcystin synthetase (*mcy*) gene cluster (Tillett *et al.*, 2000). The PKS/NRPS belongs to a family of large multi-domain enzymes that



use a modular architecture to produce important biological polyketides/nonribosomal peptides such as antibiotics. The NRPS responsible for peptide elongation is composed of at least three domains: an adenylation domain (A domain) which is responsible for substrate activation, a peptidyl carrier domain (PCP domain) that transfers the intermediate between domains and a condensation domain (C domain) which carries out peptide formation in the order C–A–PCP. PKS is composed of three domains as a basic unit: an acyltransferase domain (AT domain) responsible for substrate activation, an acyl carrier domain (ACP domain) homologous to the PCP domain and a  $\beta$ -ketoacyl synthase domain (KS domain) which catalyzes polyketide elongation assembled as KS–AT–ACP. In addition, other domains such as a carbon-methyltransferase domain (CM domain), a ketoacyl reductase domain (KR domain) and a thioesterase domain (TE domain) are required for further modification.

Comparing the modular organization of the *mcy* gene cluster with those of other PKS/NRPSs, a biosynthetic scheme for microcystin has been proposed (Tillett *et al.*, 2000). Biosynthesis of microcystin starts with McyG to assemble the side chain of the Adda residue, which is then synthesized by McyD and McyE. McyG consists of an A–PCP didomain followed by KS, AT, CM, KR and ACP domains (Fig. 1*a*). In fact, the McyG A domain accepts a wide range of phenyl-

propanoids, including *trans*-cinnamate, D-phenylalanine, *p*-coumarate, L-phenylalanine (L-Phe), phenylpyruvate, D-3-phenyllactate and L-3-phenyllactate, with various catalytic activities (Hicks *et al.*, 2006). Taking L-Phe as an example, the reaction catalyzed by McyG starts with the A domain adenylation of the carboxylate group of L-Phe with the  $\alpha$ -phosphate of ATP, forming the intermediate L-Phe-AMP (Fig. 1*b*). After adenylation, the A domain transfers L-Phe-AMP to the downstream PCP domain, which should be in the holo form (H state), in which the active-site Ser604 is post-translationally modified to contain a phosphopantetheine by 4'-phosphopantetheinyl transferase (PPTase). Subsequently, the PCP domain transfers L-Phe-AMP to the downstream KS domain, upon which the ACP domain carries a malonyl group loaded by the AT domain to the KS domain for elongation. After elongation, the ACP domain transfers the polyketide intermediate to the CM and KR domains for further modification. Lastly, the ACP domain transports the modified polyketide intermediate to the KS domain of McyD and is reloaded with a malonyl group. Notably, the A domain determines the selection and incorporation of the monomeric components into the PKS/NRPS products, whereas the PCP domain, after activation by PPTase, shuttles the intermediate from one domain to another in sequence during the polyketide/peptide synthesis.

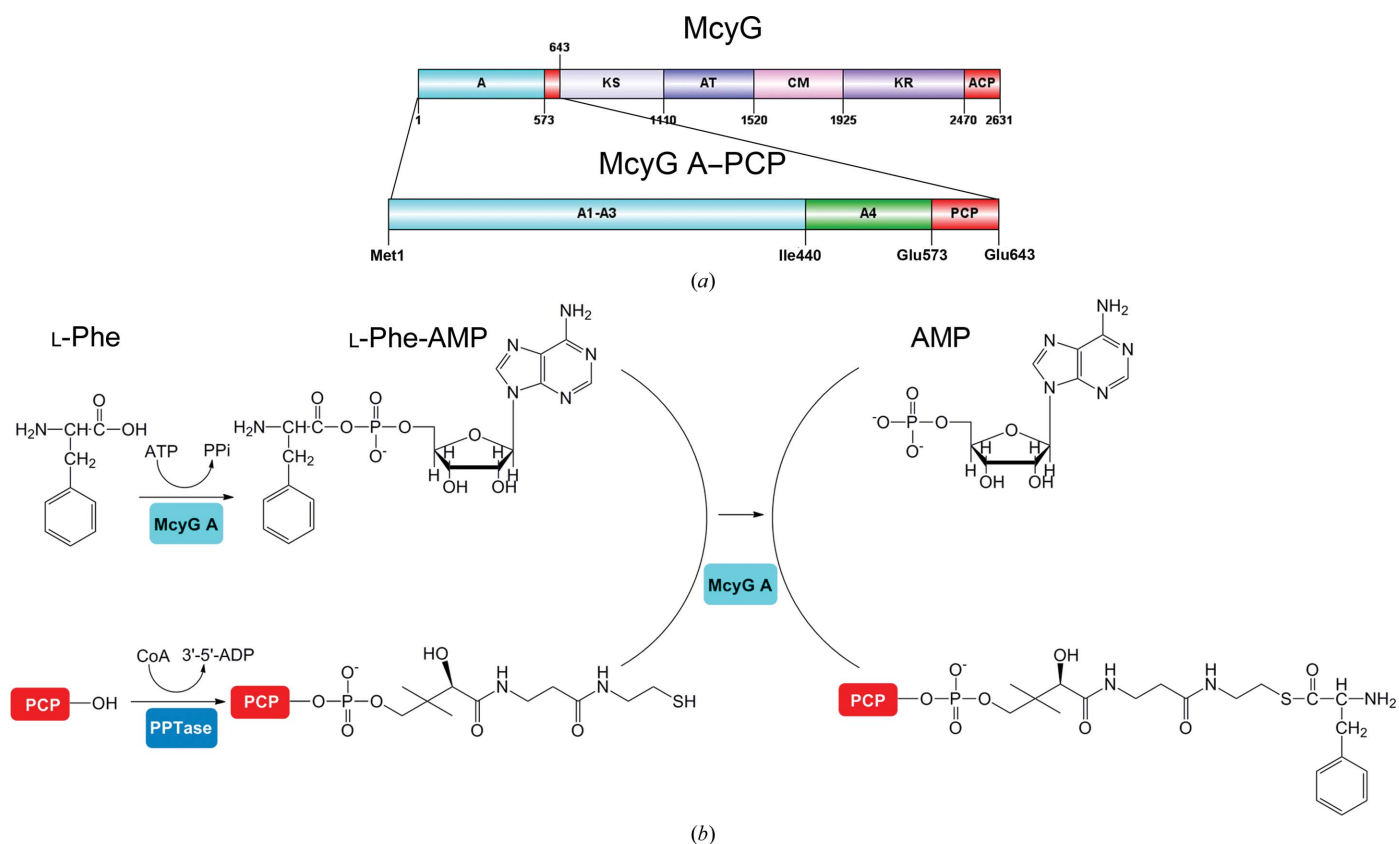


Figure 1

Organization and catalytic mechanism of McyG A–PCP. (a) Organization of McyG and McyG A–PCP. The McyG A domain is divided into four subdomains (A1–A4). (b) The McyG A domain adenylates L-Phe. Before thiolation, the PCP domain is activated by PPTase to tether the phosphopantetheinyl moiety of CoA by phosphodiester bonding to the side chain of Ser604, forming holo PCP. The intermediate L-Phe-AMP is transferred to holo PCP.

To date, two structures containing an A-PCP didomain have been deposited in the Protein Data Bank (PDB): SrfA-C (the termination module of surfactin synthetase assembled as C-A-PCP-TE; PDB entry 2vsq; Tanovic *et al.*, 2008) and PA1221 (an A-PCP didomain from *Pseudomonas aeruginosa*; PDB entry 4dg9; Mitchell *et al.*, 2012), demonstrating two different conformations of the A-PCP didomain during the catalytic cycle. In detail, the A domain of SrfA-C displays an open conformation for accepting ATP, showing the intermediate to be transferred from the downstream PCP domain to the upstream C domain (Tanovic *et al.*, 2008). In PA1221, the A domain is in a thiolated conformation guiding the holo PCP domain into the active site for thioester formation (Mitchell *et al.*, 2012). However, the mode of communication between the PCP domain and the adenylated A domain is still unknown.

Here, we have determined the crystal structure of McyG A-PCP complexed with the catalytic intermediate L-Phe-AMP. There are two molecules in one asymmetric unit, with one molecule consisting of the A-PCP didomain and the other comprising only the A domain. Detailed examination of the L-Phe-AMP binding pocket demonstrated Val227 to be a crucial residue in the selection of hydrophobic substrates. Further structural analyses identified two different interfaces displaying varied communications between the PCP domain and the A domain. We propose a catalytic cycle for adenylation and peptide transfer of the A-PCP didomain.

## 2. Materials and methods

### 2.1. Cloning, expression and purification of McyG A-PCP

The gene encoding McyG A-PCP was amplified by PCR from genomic DNA of *M. aeruginosa* PCC 7806 using primers McyG-NdeI (5'-GCACATATGATGTCTAAGCATTCATC-AGTCT-3') and McyG-XhoI (5'-TGCCTCGAGTTATTCTTCGCTTAAGAAACG-3'). The amplified fragment was cloned into pET-28b with an N-terminal hexahistidine tag. *Escherichia coli* BL21 (DE3) cells (Novagen) were transformed with the recombinant plasmid and grown in LB medium (10 g l<sup>-1</sup> NaCl, 10 g l<sup>-1</sup> Bacto tryptone, 5 g l<sup>-1</sup> yeast extract) containing 30 µg ml<sup>-1</sup> kanamycin at 37°C until an OD<sub>600 nm</sub> of 0.6 was reached. The temperature was shifted to 16°C and the cultures were induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside. After 20 h of growth, the cells were harvested and resuspended in lysis buffer (100 mM NaCl, 20 mM Tris-HCl pH 8.0). To purify the McyG A-PCP protein, resuspended cells were lysed by sonication for 30 min. After centrifugation at 12 000g for 30 min, the supernatant containing the soluble target protein was collected and loaded onto a nickel-chelating column (GE Healthcare). The binding buffer (100 mM NaCl, 20 mM Tris-HCl pH 8.0) contained 20 mM imidazole to reduce nonspecific binding of untagged proteins. The target protein was eluted with 300 mM imidazole in 6 ml binding buffer and further purified using a HiLoad 16/60 Superdex 200 column (GE Healthcare) pre-equilibrated with 100 mM NaCl, 20 mM Tris-HCl pH 8.0.

**Table 1**

Crystal parameters and data-collection and structure-refinement statistics for Se-McyG A-PCP.

Values in parentheses are for the highest resolution bin.

Data collection	
Space group	<i>P</i> 3 <sub>2</sub> 21
Unit-cell parameters (Å, °)	<i>a</i> = <i>b</i> = 120.709, <i>c</i> = 262.780, α = β = 90.00, γ = 120.00
Resolution range (Å)	54.84–2.45 (2.58–2.45)
Unique reflections	81008 (11705)
Completeness (%)	98.8 (98.7)
<i>I</i> / <i>σ</i> ( <i>I</i> )	9.6 (3.0)
<i>R</i> <sub>merge</sub> † (%)	11.8 (59.9)
Average multiplicity	5.5 (5.6)
Structure refinement	
Resolution range (Å)	54.84–2.45
<i>R</i> factor‡/ <i>R</i> <sub>free</sub> § (%)	24.8/27.3
No. of protein atoms	9475
No. of ligand atoms	68
No. of water atoms	26
R.m.s.d.¶, bond lengths (Å)	0.005
R.m.s.d., bond angles (°)	0.956
Mean <i>B</i> factors (Å <sup>2</sup> )	
Overall	43.4
Protein	43.5
Ligand	31.7
Water	32.4
Ramachandran plot††	
Most favoured (%)	97.22
Additionally allowed (%)	2.78
Outliers (%)	0
PDB entry	4r0m

† *R*<sub>merge</sub> =  $\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where *I*(*hkl*) is the intensity of an observation and  $\langle I(hkl) \rangle$  is the mean value for its unique reflection; summations are over all reflections. ‡ *R* factor =  $\sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$ , where *F*<sub>obs</sub> and *F*<sub>calc</sub> are the observed and calculated structure-factor amplitudes, respectively. § *R*<sub>free</sub> was calculated using 5% of the data, which were excluded from the refinement. ¶ Root-mean square-deviation from ideal values. †† Categories were defined by *MolProbity*.

### 2.2. Preparation of selenomethionine-labelled protein

Selenomethionine (SeMet)-labelled McyG A-PCP (Se-McyG A-PCP) was prepared for phase determination. Se-McyG A-PCP was obtained by expressing McyG A-PCP in *E. coli* strain B834 (DE3) (Novagen). The transformed cells were inoculated into LB medium and grown at 37°C until an OD<sub>600 nm</sub> of 0.2 was reached. The cells were harvested and then washed twice with M9 medium. The cells were then cultured in SeMet medium (M9 medium with 25 mg l<sup>-1</sup> L-SeMet and the other essential amino acids at 50 mg l<sup>-1</sup>) to an OD<sub>600 nm</sub> of 0.6 and induced using 0.2 mM isopropyl β-D-1-thiogalactopyranoside. The remaining steps of protein expression and purification were identical to those used for the native McyG A-PCP except that all buffers contained 5% (v/v) glycerol and 14 mM β-mercaptoethanol. Se-McyG A-PCP was concentrated to a final concentration of 58 mg ml<sup>-1</sup> for crystallization. The purity was estimated using 10% SDS-PAGE. The protein concentration was determined by *A*<sub>280 nm</sub> using an extinction coefficient of 75 860 M<sup>-1</sup> cm<sup>-1</sup>.

### 2.3. Crystallization, data collection and processing

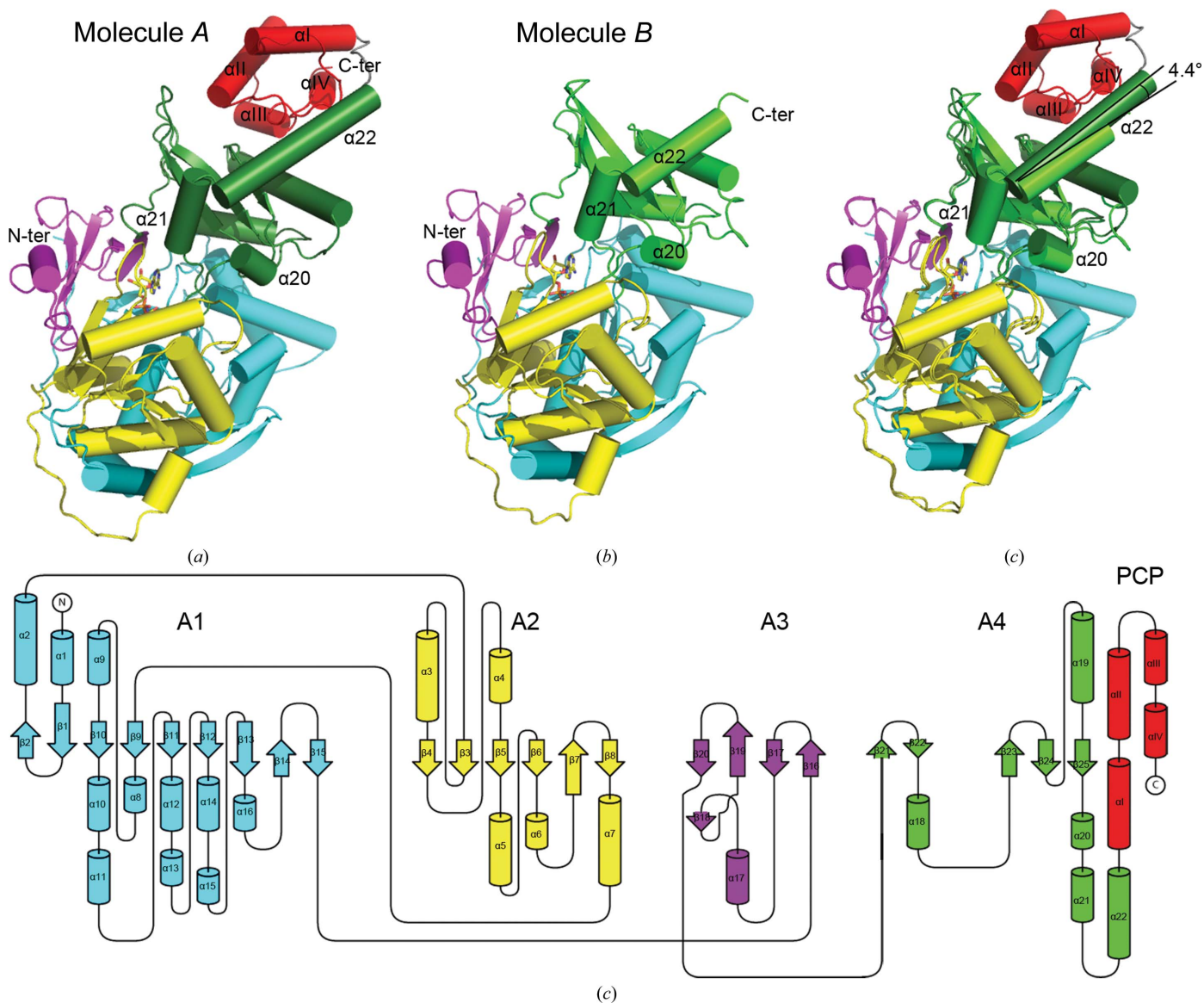
Before crystallization, Se-McyG A-PCP was incubated with 5 mM L-Phe, 5 mM ATP, 1 mM MgCl<sub>2</sub> for 30 min at 25°C. Initial crystallization screening for Se-McyG A-PCP was performed using the hanging-drop vapour-diffusion method at

16°C in 96-well plates. The crystallization screening kits comprised Crystal Screen, Crystal Screen 2, Index, Grid Screen, SaltRx (Hampton Research) and Buffer and pH screen (Radaev *et al.*, 2006). Se-McyG A-PCP crystals were obtained in a condition consisting of 1.8 M ammonium sulfate, 0.1 M NaKHPO<sub>4</sub> pH 6.5. During optimization in 24-well plates, the hanging drop consisted of 1 µl protein sample and 1 µl reservoir solution consisting of 1.2–2.2 M ammonium sulfate, 0.1 M NaKHPO<sub>4</sub> pH 6.5. Crystals appeared in one week and grew to maximum size over three weeks. Se-McyG A-PCP crystals were briefly soaked in reservoir solution supplemented with 30% glycerol as a cryoprotectant, mounted on loops and flash-cooled in liquid nitrogen. SeMet-derivative data were collected from single crystals at 100 K in a liquid-nitrogen stream on beamline 17U at the Shanghai Synchrotron Radiation Facility (SSRF) using an ADSC Q315r CCD

detector (MAR Research, Germany). All diffraction data were indexed, integrated and scaled with *iMosflm* (Battye *et al.*, 2011).

#### 2.4. Structure determination and refinement

The crystal structure of McyG A-PCP was determined using the single-wavelength anomalous dispersion phasing method from a single SeMet-substituted protein crystal to a maximum resolution of 2.45 Å. The *AutoSol* program from *PHENIX* (Adams *et al.*, 2002) was used to locate the heavy atoms, and the phases were calculated and further improved using partial model extension with *OASIS* (Hao *et al.*, 2000). Automatic model building was carried out using *AutoBuild* in *PHENIX*. Refinement was carried out using the maximum-likelihood method implemented in *REFMAC5* (Murshudov *et*



**Figure 2** Overall structure of McyG A-PCP. Two molecules in one asymmetric unit are shown: (a) molecule A and (b) molecule B. The A1–A3 subdomains of the McyG A domain are shown in cyan, yellow and magenta, respectively, whereas the A4 subdomains in molecules A and B are shown in dark green and light green, respectively. The PCP domain is coloured red. (c) Comparison of molecules A and B. (d) Topological diagram of molecule A.



*al.*, 2011) as part of the *CCP4* program suite (Winn *et al.*, 2011) and rebuilt interactively using *Coot* (Emsley & Cowtan, 2004). The final model was evaluated with *MolProbity* (Chen *et al.*, 2010) and *PROCHECK* (Laskowski *et al.*, 1993). Crystallographic parameters and data-collection statistics are listed in Table 1. All figures showing the structure were prepared with *PyMOL* (DeLano, 2002).

### 3. Results and discussion

#### 3.1. Overall structure of McyG A-PCP

The structure of McyG A-PCP was determined at 2.45 Å resolution in space group  $P3_221$ , with two monomers (namely molecules *A* and *B*) in one asymmetric unit. Molecule *A* consists of an A domain (Ser5–Ser572) and a PCP domain (Glu573–Ser641) (Fig. 2*a*). However, molecule *B* comprises only an A domain corresponding to residues Ser7–Lys561 (Fig. 2*b*) owing to the electron density of residues Glu562–Ser643 being disordered. The A domain of both molecules is bound to one molecule of L-Phe-AMP.

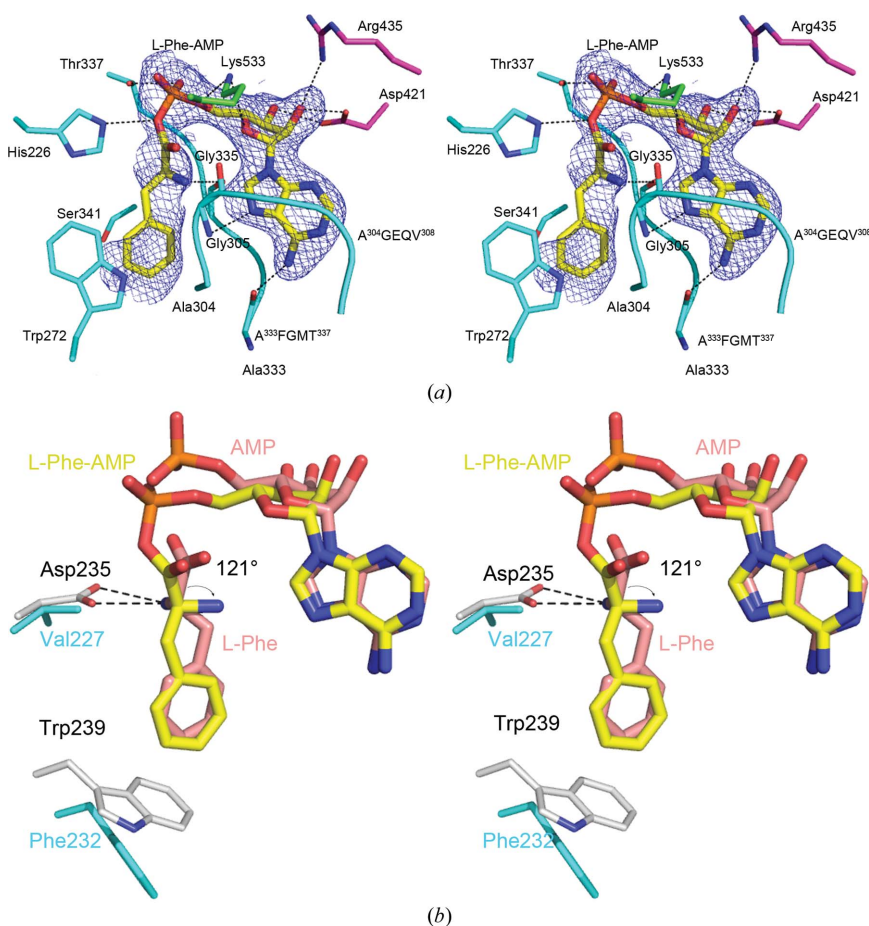
The McyG A domain is composed of two domains: the large domain (A1–A3 subdomains; Ser5–Gly434) and the small domain (A4 subdomain; Ile440–Lys566) connected by a flexible loop (Arg435–Glu439) which permits the A4 subdomain to adopt different orientations with respect to the A1–A3 subdomains. The A1 subdomain consists of a nine-stranded  $\beta$ -sheet ( $\beta_1$ ,  $\beta_2$ ,  $\beta_9$ – $\beta_{15}$ ) surrounded by 11  $\alpha$ -helices ( $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_8$ – $\alpha_{16}$ ), forming a distorted  $\beta$ -barrel. The A2 subdomain is composed of a six-stranded  $\beta$ -sheet ( $\beta_3$ – $\beta_8$ ) flanked by five  $\alpha$ -helices ( $\alpha_3$ – $\alpha_7$ ), whereas the A3 subdomain contains a distorted five-stranded  $\beta$ -sheet ( $\beta_{16}$ – $\beta_{20}$ ) in addition to one  $\alpha$ -helix ( $\alpha_{17}$ ). Collectively, the A1–A3 subdomains display a distorted  $\beta$ -barrel and two  $\beta$ -sheets which pack against each other to form a five-layered  $\alpha\beta\alpha\beta\alpha$  domain structure. The A4 subdomain begins with an antiparallel  $\beta$ -sheet containing two small  $\beta$ -strands ( $\beta_{21}$  and  $\beta_{22}$ ) followed by a central three-stranded  $\beta$ -sheet ( $\beta_{23}$ – $\beta_{25}$ ) which is surrounded by five  $\alpha$ -helices ( $\alpha_{18}$ – $\alpha_{22}$ ). Notably, the structure of the McyG A domain displays an adenylated conformation, binding to one molecule of L-Phe-AMP.

As is well known, three states of the PCP domain have been observed according to the variation of four  $\alpha$ -helices ( $\alpha I$ – $\alpha IV$ ): an unmodified state (A state), a modified state (H state) and an intermediary state between the A state and the H state (the A/H state) (Koglin *et al.*, 2006). Specifically, in the A

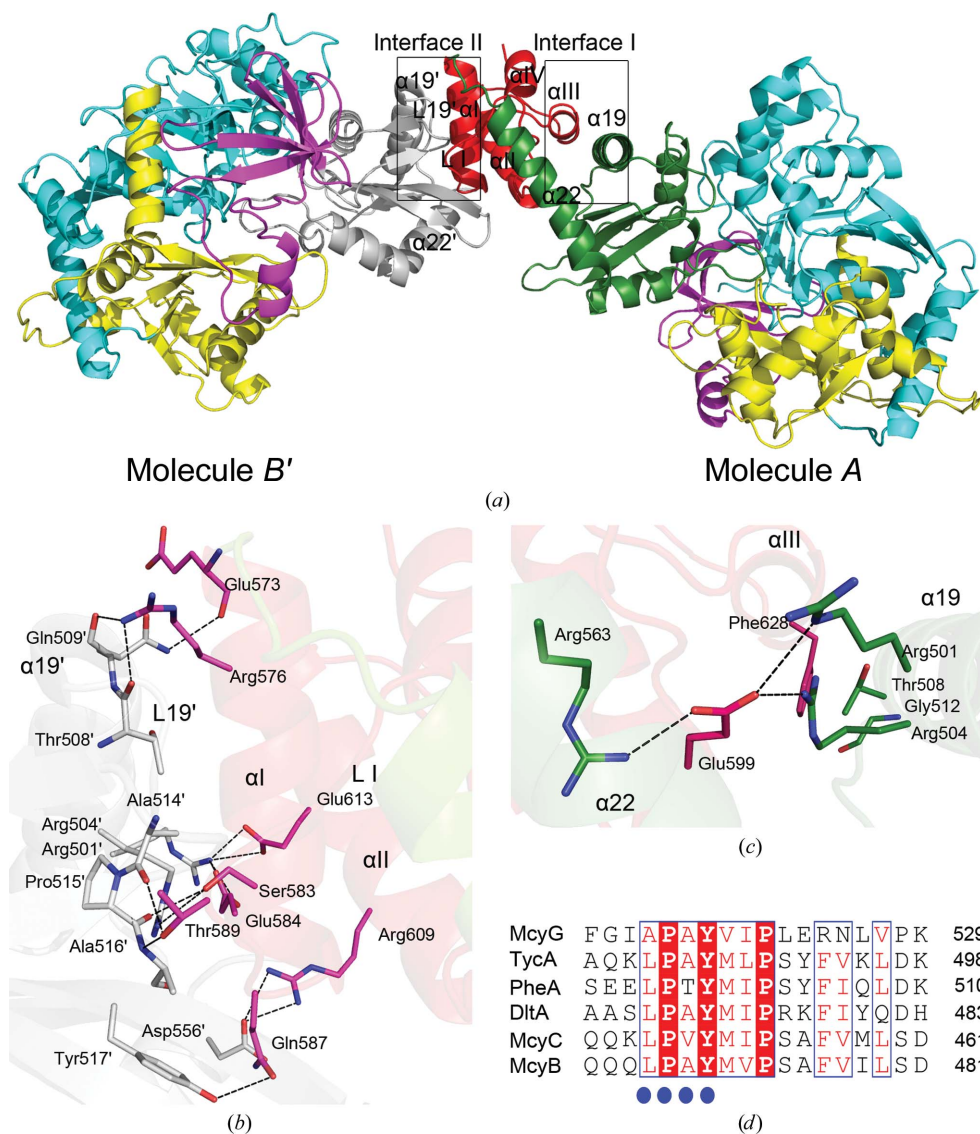
state  $\alpha III$  changes to a loop and the other three helices are shorter, in the H state  $\alpha III$  is also disordered and helix  $\alpha IV$  moves parallel to  $\alpha II$ , while the A/H state is composed of four  $\alpha$ -helices named  $\alpha I$ – $\alpha IV$ . In McyG A-PCP, the PCP domain of molecule *A* consists of four  $\alpha$ -helices ( $\alpha I$ – $\alpha IV$ ) which were well defined in the electron-density map in the conformation of the A/H state with the active-site Ser604 unmodified.

#### 3.2. Comparison of the two A-PCP didomains in each asymmetric unit

Superposition of the two A-PCP molecules (molecules *A* and *B*) in each asymmetric unit yielded root-mean-square deviations (r.m.s.d.s) of 0.466 and 0.901 Å over 427  $C^\alpha$  atoms of the A1–A3 subdomains and 114  $C^\alpha$  atoms of the A4 subdomain, respectively. It is notable that three  $\alpha$ -helices ( $\alpha_{20}$ – $\alpha_{22}$ ) of the A4 subdomain slant, with helix  $\alpha_{22}$  flipped by 4.4° (Fig. 2*c*). In the structure of SrfA-C, the segment corresponding to helix  $\alpha_{22}$  adopts a loop conformation to enable the PCP domain to approach the upstream C domain,



**Figure 3**  
L-Phe binding pocket of the McyG A domain and comparison of the pocket with that of PheA. (a) A stereoview of the L-Phe-AMP binding pocket of the McyG A domain. L-Phe-AMP is superimposed with the  $F_o - F_c$  map contoured at  $3\sigma$ . Residues interacting with L-Phe-AMP are shown in stick representation. Polar interactions are indicated by dashed lines. (b) Comparison of the substrate-binding pockets of the McyG A domain and PheA. L-Phe-AMP and residues from the McyG A domain are shown in yellow and cyan, respectively. L-Phe and AMP are shown in pink. Residues from PheA are coloured grey.



**Figure 4** Interfaces between the A4 subdomain and the PCP domain. (a) Interface I is formed between the A4 subdomain (green) and the PCP domain (red) within molecule A. Interface II is formed between the PCP domain of molecule A and the A4 subdomain (grey) of molecule B' from the neighbouring asymmetric unit. (b, c) Residues participating in stabilizing (b) interface I and (c) interface II. (d) Multiple sequence alignment of the McyG A domain and other adenylation domains. Conserved residues involved in the formation of interface II are indicated by blue dots. The multiple sequence alignment was performed using *ClustalW2* (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) and *ESPrpt* (<http://esprpt.ibcp.fr/ESPrpt/cgi-bin/ESPrpt.cgi>). All sequences were downloaded from the UniProtKB database and the Protein Data Bank. McyG, UniProt ID A8YJW1; McyB, A8YJV6; McyC, Q9RNA9; TycA (tyrocidine synthetase 1), P09095; Phe (the phenylalanine-activating subunit of gramicidin synthetase 1), PDB entry 1amu (Conti *et al.*, 1997); DltA (D-alanyl carrier protein ligase), PDB entry 3e7w (Yonus *et al.*, 2008).

whereas the counterpart in PA1221 is more flexible, which facilitates interaction between the PCP domain and the A domain (Tanovic *et al.*, 2008; Mitchell *et al.*, 2012). Therefore, we speculated that the slanting of the  $\alpha 20$ – $\alpha 22$  helices in addition to the high flexibility of the segment corresponding to helix  $\alpha 22$  is involved in the shuttling and/or conformational changes of the PCP domain during the catalytic cycle.

### 3.3. The L-Phe-AMP binding site

Taking molecule B as an example, the L-Phe-AMP binding pocket is positioned at the interface between the large and

small domains. From a closer viewpoint, the L-Phe-AMP is mainly surrounded by residues from the A1 subdomain and is covered by the A4 subdomain like a lid. In detail, L-Phe-AMP lies in a slot sandwiched between two loops consisting of A<sup>304</sup>GEQV<sup>308</sup> and A<sup>333</sup>FGMT<sup>337</sup>, respectively (Fig. 3a). With respect to the AMP moiety of L-Phe-AMP, the adenine base is fixed by two hydrogen bonds, one of which is formed between the N6 amino group of the adenine base and the  $\alpha$ -carbonyl of Ala333 and the other of which is formed between the N7 amino group and the  $\alpha$ -carbonyl of Gly305. The ribose 2' and 3' hydroxyls interact with Asp421 and Arg435 from the A3 subdomain and the ribose 4' and 5' O atoms bind to the side chain of residue Lys533. The  $\alpha$ -phosphate interacts with the side chain of Thr337. In addition, His226 forms a hydrogen bond to the bridging O atom between the L-Phe group and the AMP part. Moreover, the  $\alpha$ -amino group of the L-Phe group is fixed by the formation of one hydrogen bond to the  $\alpha$ -carbonyl of Gly335. The benzene ring of the L-Phe group is stabilized by hydrophobic interactions *via* being sandwiched between the side chain of Trp272 and the main chains of Ala333, Gly335 and Ser341.

To reveal further characteristics of the substrate-binding pocket of McyG A–PCP, superposition with PheA, which is a classic A domain activating L-Phe (Conti *et al.*, 1997), was performed. Generally, residues that participate in fixing the AMP part of McyG A–PCP and PheA are highly conserved (data not shown), whereas two notable features among the L-Phe group binding residues were observed in McyG A–PCP. Val227 in the McyG A domain is positioned at the same position as the corresponding residue Asp235 of PheA (Fig. 3b). It is known that Asp235 is conserved in all amino-acid adenylation domains, participating in the formation of hydrogen bonds to the  $\alpha$ -amino group of the amino-acid substrate (Stachelhaus *et al.*, 1999). Remarkably, the side chain of Val227 provides a hydrophobic environment which rotates

the  $\alpha$ -amino group of L-Phe-AMP in the McyG A domain through  $121^\circ$  away from the counterpart  $\alpha$ -amino group of L-Phe in PheA. Moreover, an ATP-PP<sub>i</sub> exchange assay of McyG A-PCP towards various substrates identified two hydrophobic substrates lacking  $\alpha$ -amino groups as displaying the highest activities: *trans*-cinnamate has the highest activity, while that of hydrocinnamate was lower (Hicks *et al.*, 2006). In fact, compared with wild-type McyG A-PCP, V227T and V227D mutations resulted in an increased  $K_m$  value towards *trans*-cinnamate by approximately threefold and 23-fold and a  $k_{cat}/K_m$  value that was decreased by 43 and 96.6%, respectively (Supplementary Table S1). Therefore, we speculated that Val227 in McyG is most likely to be critical for the selection of hydrophobic substrates. In PheA, the side chain of Trp239 is positioned at the bottom as a lid, covering the pocket with an average distance of 3.6 Å from the benzene ring of L-Phe. However, the corresponding distance from the benzene

ring of L-Phe to the side chain of Phe232 in McyG A is as long as 5 Å, leading to a deeper substrate-binding pocket which may accommodate substrates that are longer than L-Phe to some degree. Indeed, the longer substrate hydroxycinnamate can be recognized by the McyG A domain (Hicks *et al.*, 2006).

### 3.4. Interfaces between the A domain and the PCP domain

We observed two different interfaces between the A domain and the PCP domain: interface I of 641 Å<sup>2</sup> buried area is formed between the A domain and the PCP domain within molecule A, while interface II of 952 Å<sup>2</sup> is formed between the PCP domain of molecule A and the A domain of molecule B' from the neighbouring asymmetric unit (Fig. 4a). In interface I, hydrophobic interactions are formed between Gly512 ( $\alpha$ 19), Thr508 ( $\alpha$ 19) and Phe628 ( $\alpha$ III). In addition, Glu599 from loop I (LI) is stabilized by forming hydrogen bonds to Arg501 ( $\alpha$ 19), Arg504 ( $\alpha$ 19) and Arg563 ( $\alpha$ 22) (Fig. 4b). It is known that in the structure of the TycC\_PCP (S45A)-Sfp complex (the PCP domain of tyrocidine synthetase 3 complexed with the PPTase Sfp), Gln40, which corresponds to Glu599 of McyG A-PCP, is exposed to participate in the interaction between the PCP domain and the PPTase Sfp (Tufar *et al.*, 2014). Additionally, in EntB\_PCP (the PCP domain of enterobactin synthetase B), Ala268, which corresponds to Phe628 in McyG A-PCP, is crucial for interaction with the downstream C domain (Lai *et al.*, 2006). Clearly, both Glu599 and Phe628 are hindered from participating in forming an interface between the PCP domain and the downstream domains or PPTase in interface I. Moreover, the active-site Ser604 orients towards the A domain and is buried in interface I; thus, the PCP domain in molecule A adopts a blocked conformation which is prevented from interacting with downstream domains and PPTase.

Interface II is stabilized by a network of hydrogen bonds and salt bridges (Fig. 4c). The hydrogen bonds are formed as follows: from the main chain of Glu573 ( $\alpha$ I) to the side chain of Gln509' ( $\alpha$ 19'), from the side chain of Arg576 ( $\alpha$ I) to the  $\alpha$ -carbonyl of Thr508' ( $\alpha$ 19') and

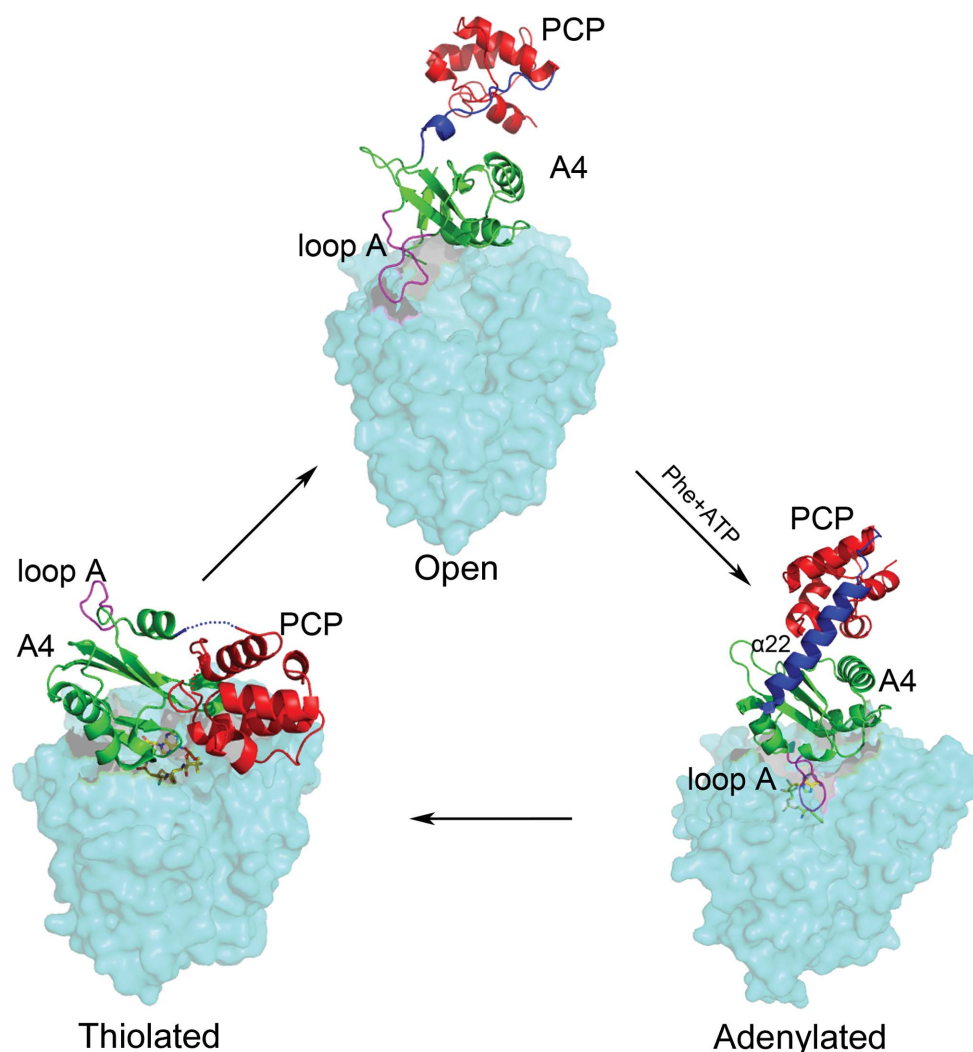


Figure 5

A putative catalytic cycle of the A-PCP didomain. Three states, open (SrfA-C), adenylylated (McyG) and thiolated (PA1221), are shown to represent the reaction cycle of the A-PCP didomain. The A1-A3 subdomains are shown in cyan in the same orientation in the three states and loop A is shown in magenta. The A4 subdomain and the PCP domain are shown in green and red, respectively. Helix  $\alpha$ 22 of adenylylated McyG and the corresponding segments of SrfA-C and PA1221 are shown in blue.



Gln509' ( $\alpha 19'$ ), from the side chain of Ser583 ( $\alpha 1$ ) to the side chain of Arg504' ( $\alpha 19'$ ) and the  $\alpha$ -carbonyl of Pro515' (loop19'; L19'), from Gln587 (LI) to Tyr517' (L19') and from the side chain of Thr589 (LI) to the  $\alpha$ -carbonyl of Ala514' (L19') and Ala516' (L19'). In addition, the NH1 and NH2 amino groups of Arg609 ( $\alpha 22'$ ) form two salt bridges to the side chain of Asp556' ( $\alpha 22'$ ). The side chains of Glu584 ( $\alpha 1$ ) and Glu613 ( $\alpha 22'$ ) form four salt bridges to Arg501'. Biochemical analyses of the adenylation reaction of TycA-A (the A domain of tyrocidine synthetase 1) demonstrated that residues Pro485 and Tyr487, which correspond to Pro515' and Tyr517' from L19' of McyG A-PCP, are crucial in transferring the intermediate to the PCP domain (Bučević-Popović *et al.*, 2012). Multiple sequence alignment showed that Pro515 and Tyr517 of McyG A-PCP are conserved among the adenylation domains, suggesting that the interaction between L19' and the PCP domain may be related to intermolecular recognition and intermediate transfer between the A domain and the PCP domain (Fig. 4d).

### 3.5. A putative catalytic cycle of the A-PCP didomain

We observed the intact A-PCP didomain of McyG in the adenylation conformation in molecule A. Structural comparison of McyG A-PCP with SrfA-C and PA1221 showed that the conformation of the A1-A3 subdomains varies little, whereas both the A4 subdomain and the PCP domain adopt changing conformations and move in relation to the A1-A3 subdomains during the catalytic reaction. Specifically, in adenylation McyG A-PCP loop A (Pro527-Gln535) moves to cover the substrate-binding pocket, with Lys533 hydrogen-bonding to the intermediate L-Phe-AMP, whereas the corresponding loop in SrfA-C A-PCP is located far away, resulting in a more open pocket. Moreover, in adenylation McyG helix  $\alpha 22$  is well ordered and the PCP domain is more close to the A4 subdomain, with the active-site Ser604 blocked to hinder the PCP domain from interacting with the downstream KS domain or A domain. However, in SrfA-C A-PCP the segment corresponding to helix  $\alpha 22$  of McyG transforms into a flexible loop and the PCP domain is positioned far away from the A4 subdomain, with the active-site serine approaching the upstream C domain. Further structural analyses of McyG A-PCP and PA1221 revealed that the A4 subdomain of McyG A-PCP needs to rotate  $140^\circ$  to enable the holo PCP to interact with the intermediate, along with the rearrangement of helix  $\alpha 22$  and the movement of the PCP domain to form the thiolated conformation.

It is known that in three-dimensional cryo-microscopy structures of holo PikAIII (module 5 of pikromycin PKS with a modified ACP domain organized as KS-AT-KR-holo ACP<sub>5</sub>), pentaketide-KS<sub>5</sub>-PikAIII and  $\beta$ -keto-hexaketide-PikAIII, the ACP domain, AT domain and KS domain undergo conformational rearrangements during the enzymatic cycle. In detail, once the ACP<sub>5</sub> domain has transported the  $\beta$ -keto-hexaketide to the KR domain for modification, the AT domain moves to cover the side entrance of the KS domain in order to prevent the KS domain from accepting the inter-

mediate transported by the upstream ACP<sub>4</sub> domain and to keep the catalytic cycle in sequence (Whicher *et al.*, 2014; Dutta *et al.*, 2014). The adenylation of the A-PCP didomain in our structure prevents the PCP domain from interacting with the downstream KS domain and PPTase until the downstream ACP domain has finished transporting the intermediate to the KS domain of McyD and has reloaded the malonyl group. Thus, we captured a transient conformation of the A-PCP didomain that induces the downstream domains to react in sequence.

Based on these findings, we propose a putative catalytic cycle for McyG A-PCP. Upon adenylation, the A-PCP didomain is blocked in a conformation in which the PCP domain is prevented from interacting with the downstream domains or PPTase. After reloading of the malonyl group onto the downstream ACP domain by the AT domain, the A4 subdomain rotates by  $140^\circ$  and helix  $\alpha 22$  rearranges to help the PCP domain attach to the intermediate, forming the thiolated conformation. Finally, the A-PCP didomain converts into an open conformation to allow the PCP domain to more easily transport the intermediate to the downstream KS domain for elongation and to accept substrates for next reaction cycle (Fig. 5).

### Acknowledgements

We thank the beamline staff of the Shanghai Synchrotron Radiation Facility (SSRF) for technical help during X-ray data collection. This work was supported by the Chinese National Natural Science Foundation (Grant Nos. 31370757 and 31070652), a Scientific Research Grant from Hefei Science Center of CAS and the Ministry of Education of China (Grant No. 20133402110023) and the Program for Changjiang Scholars and Innovative Research Team in University.

### References

- 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. & Terwilliger, T. C. (2002). *Acta Cryst.* **D58**, 1948–1954.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Bučević-Popović, V., Sprung, M., Soldo, B. & Pavela-Vrančić, M. (2012). *Chembiochem*, **13**, 1913–1920.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. (1997). *EMBO J.* **16**, 4174–4183.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.
- Dittmann, E., Neilan, B. A., Erhard, M., von Döhren, H. & Börner, T. (1997). *Mol. Microbiol.* **26**, 779–787.
- Dutta, S., Whicher, J. R., Hansen, D. A., Hale, W. A., Chemler, J. A., Congdon, G. R., Narayan, A. R., Håkansson, K., Sherman, D. H., Smith, J. L. & Skiniotis, G. (2014). *Nature (London)*, **510**, 512–517.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Goldberg, J., Huang, H.-B., Kwon, Y.-G., Greengard, P., Nairn, A. C. & Kuriyan, J. (1995). *Nature (London)*, **376**, 745–753.
- Hao, Q., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2000). *J. Appl. Cryst.* **33**, 980–981.
- Hicks, L. M., Moffitt, M. C., Beer, L. L., Moore, B. S. & Kelleher, N. L. (2006). *ACS Chem. Biol.* **1**, 93–102.



- Honkanen, R. E., Zwiller, J., Moore, R. E., Daily, S. L., Khatra, B. S., Dukelow, M. & Boynton, A. L. (1990). *J. Biol. Chem.* **265**, 19401–19404.
- Jochimsen, E. M., Carmichael, W. W., An, J. S., Cardo, D. M., Cookson, S. T., Holmes, C. E. M., Antunes, M. B., de Melo Filho, D. A., Lyra, T. M., Barreto, V. S. T., Azevedo, S. M. F. O. & Jarvis, W. R. (1998). *N. Engl. J. Med.* **338**, 873–878.
- Koglin, A., Mofid, M. R., Löhr, F., Schäfer, B., Rogov, V. V., Blum, M. M., Mittag, T., Marahiel, M. A., Bernhard, F. & Dötsch, V. (2006). *Science*, **312**, 273–276.
- Lai, J. R., Koglin, A. & Walsh, C. T. (2006). *Biochemistry*, **45**, 14869–14879.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Mitchell, C. A., Shi, C., Aldrich, C. C. & Gulick, A. M. (2012). *Biochemistry*, **51**, 3252–3263.
- Moore, R. E., Chen, J. L., Moore, B. S., Patterson, G. M. L. & Carmichael, W. W. (1991). *J. Am. Chem. Soc.* **113**, 5083–5084.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst. D* **67**, 355–367.
- Nishiwaki-Matsushima, R., Ohta, T., Nishiwaki, S., Suganuma, M., Kohyama, K., Ishikawa, T., Carmichael, W. W. & Fujiki, H. (1992). *J. Cancer Res. Clin. Oncol.* **118**, 420–424.
- Pouria, S., de Andrade, A., Barbosa, J., Cavalcanti, R. L., Barreto, V. T., Ward, C. J., Preiser, W., Poon, G. K., Neild, G. H. & Codd, G. A. (1998). *Lancet*, **352**, 21–26.
- Radaev, S., Li, S. & Sun, P. D. (2006). *Acta Cryst.* **D62**, 605–612.
- Rinehart, K. L., Namikoshi, M. & Choi, B. W. (1994). *J. Appl. Phycol.* **6**, 159–176.
- Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. (1999). *Chem. Biol.* **6**, 493–505.
- Tanovic, A., Samel, S. A., Essen, L.-O. & Marahiel, M. A. (2008). *Science*, **321**, 659–663.
- Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T. & Neilan, B. A. (2000). *Chem. Biol.* **7**, 753–764.
- Tufar, P., Rahighi, S., Kraas, F. I., Kirchner, D. K., Löhr, F., Henrich, E., Köpke, J., Dikic, I., Güntert, P., Marahiel, M. A. & Dötsch, V. (2014). *Chem. Biol.* **21**, 552–562.
- Whicher, J. R., Dutta, S., Hansen, D. A., Hale, W. A., Chemler, J. A., Dosey, A. M., Narayan, A. R., Håkansson, K., Sherman, D. H., Smith, J. L. & Skiniotis, G. (2014). *Nature (London)*, **510**, 560–564.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Yonus, H., Neumann, P., Zimmermann, S., May, J. J., Marahiel, M. A. & Stubbs, M. T. (2008). *J. Biol. Chem.* **283**, 32484–32491.