Crystallographic analysis of a cupin superfamily enzyme from *Microcystis aeruginosa* involved in aeruginosin biosynthesis

Xiaoting Qiu,^{a*} Cui Xu^{a,b} and Xu Chen^a

Received 21 October 2015

Accepted 17 November 2015

Edited by R. A. Pauptit, Macclesfield, England

Keywords: cupin domain; AerE; 2-carboxy-6-hydroxyoctahydroindole moiety; aeruginosin.

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

^aKey Laboratory of Applied Marine Biotechnology of Ministry of Education, School of Marine Sciences, Ningbo University, Ningbo, Zhejiang 315211, People's Republic of China, and ^bCollege of Science and Technology, Ningbo University, Ningbo, Zhejiang 315211, People's Republic of China. *Correspondence e-mail: xiaotingqiu@126.com

Aeruginosins are a class of cyanobacteria-derived bioactive linear tetrapeptides composed of nonproteinogenic amino-acid residues, such as the 2-carboxy-6-hydroxyoctahydroindole (Choi) moiety, which is the hallmark of aeruginosin. The biosynthetic pathway of the Choi moiety remains elusive. Previous studies have suggested that AerE, a protein that possesses two cupin domains, participates in the biosynthesis of the Choi moiety. In this study, recombinant AerE from *Microcystis aeruginosa*, which was overexpressed in *Escherichia coli* and purified by Ni²⁺-chelating affinity and gel-filtration chromatography, was successfully crystallized and X-ray diffraction analysis was performed. The crystal belonged to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters $a = 34.770$, $b = 62.133$, $c = 87.401$ Å. The diffraction data from the crystal were scaled to a maximum resolution of 1.60 Å. The calculated Matthews coefficient of the crystal is $1.93 \text{ Å}^3 \text{ Da}^{-1}$, suggesting that there is one molecule in the asymmetric unit.

1. Introduction

Secondary metabolites that demonstrate diverse biological activities are widely distributed in cyanobacteria. They are a source of drug candidates such as cyclic and linear peptide toxins (Welker & von Döhren, 2006; Welker *et al.*, 2006). Studies of the biosynthetic mechanism of the nonproteinogenic amino acid present in the cyanobacteria-derived peptide toxins can facilitate the biomimetic synthesis of these bioactive compounds. Aeruginosins are a class of cyanobacteria-derived linear tetrapeptides (Ishida *et al.*, 2007), with their hallmark being the existence of a 2-carboxy-6-hydroxyoctahydroindole (Choi) moiety, a bicyclic nonproteinogenic amino-acid residue which is essential for the function of aeruginosin in inhibiting various types of serine-type proteases (Sandler *et al.*, 1998; Rios Steiner *et al.*, 1998). The biosynthetic pathway of Choi has been deduced to be initiated by the decarboxylation of prephenate (Mahlstedt *et al.*, 2010), while the subsequent steps involved in Choi biosynthesis have yet to be elucidated.

Previous studies demonstrated that *aerE*, a gene located in the *aer* gene cluster, is involved in Choi biosynthesis (Ishida *et al.*, 2009). However, the catalytic specificity and the catalytic mechanism of AerE, the product of the *aerE* gene, have not been unravelled. Analysis of the sequence motif shows that there are two cupin domains in AerE (in AerE from *Microcystis aeruginosa* the first cupin domain starts at residue 9 and ends at residue 114 and the second cupin domain starts at residue 116 and ends at residue 207). The cupin domain is a double-stranded β -helix domain which is widely distributed

in various types of metal-dependent enzymes possessing different catalytic activities (Anantharaman *et al.*, 2003). The enzymes containing cupin domains can be assigned to the cupin superfamily as they exhibit identical folding and a similar conformation of the active site. Variation of the residues constituting the active site and the identity of the bound metal ion lead to the different catalytic specificities of the members of the cupin superfamily (Dunwell *et al.*, 2001).

BacB and YxaG from *Bacillus subtilis*, which are sequence homologues of AerE, are two enzymes in the cupin superfamily that have been structurally and functionally analyzed. BacB was demonstrated to be an oxidase that catalyzes the second step of the biosynthetic pathway from prephenate to the anticapsin moiety of bacilysin (Rajavel *et al.*, 2009), and it can also facilitate the isomerization of the decarboxylation product of prephenate (Mahlstedt *et al.*, 2010). YxaG is an Fe²⁺-containing quercetin 2,3-dioxygenase that exhibits multiple functions, and its activity can be regulated by varying the metal ion in the active site (Gopal *et al.*, 2005). Both BacB and YxaG have two cupin domains, and each domain coordinates a metal ion, which is a critical cofactor for enzymatic activity. Sequence alignment (Fig. 1) illustrates that the metal ion-coordinating residues in the N-terminal cupin domains of BacB and YxaG are conserved in AerE, while the corresponding residues in the C-terminal cupin domain are only partially conserved in AerE, suggesting that the C-terminal cupin domain of AerE is probably not capable of binding

metal ions. Moreover, several key residues in the N-terminal cupin domain of BacB that participate in substrate recognition and orientation are not conserved in AerE. These results imply that AerE presumably possesses a similar catalytic specificity to these bicupin enzymes, while possessing a different substrate-binding mode and catalytic mechanism. Here, we chose AerE from *M. aeruginosa* (Ma-AerE) as a target, which was expressed in *E. coli*, purified and successfully crystallized, and X-ray diffraction data were collected from the crystal. A preliminary foundation for further structural and enzymatic studies of the role of AerE in Choi biosynthesis has thus been established.

2. Materials and methods

2.1. Cloning, expression and purification

The coding region for full-length Ma-AerE was PCR-amplified from the genomic DNA of *M. aeruginosa* (strain PCC7806) and then inserted into the expression plasmid pET-22b (Novagen) between the NdeI and XhoI restriction sites. The cloning sequence was confirmed by sequencing (Invitrogen). The plasmid was transformed into *E. coli* BL21 (DE3) cells. Transformed bacteria were cultured at 310 K to an OD_{600 nm} of 0.8 in 1.6 l LB medium containing 50 µg ml⁻¹ ampicillin. The cells were then induced for a further 24 h with 0.25 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at

Figure 1

Multiple sequence alignment of Ma-AerE with BacB and YxaG from *B. subtilis*. The metal-coordinating residues in the N-terminal and C-terminal cupin domains and the key residues involved in substrate recognition and orientation in BacB are indicated by blue, green and red stars, respectively. The alignment was generated using the *ClustalW* web server and the figure was prepared by *ESPrpt*.

Table 1
Macromolecule-production information.

Source organism and DNA source	<i>M. aeruginosa</i> (strain PCC7806)
Forward primer†	GGAATTCATATGCTCTGAATTTTCCAGTTC
Reverse primer‡	CCGCTCGAGAAATTAATTTTAAATCAATTAATA-TCTTG
Cloning vector and expression vector	pET-22b
Expression host	<i>E. coli</i> BL21 (DE3)
Complete amino-acid sequence of the construct produced§	MSEFFPVPKPIKLNPHVELEVFCQDTIFQLSVI- APNAKLESHQHPESQIGMVLSGELELYIKDVI- KPLRALQDIHVADANVSHGFVNPLSEPMIGFD- LKRTSSLPSEDVVLTLNNDKITHLPCQSV- KGSWFELVMMKIPSGYSIPPHQGEQEIGFIL- NGKLEIFIENEQCLEYGQIYYAPSKVLKKGYN- SSNQDINLIKILILEHHHHH

† The *Nde*I site is underlined. ‡ The *Xho*I site is underlined. § The cloning artifact is underlined.

289 K for protein expression. The cell pellet, which was collected by centrifugation at 9990g for 5 min at 281 K, was suspended in 50 ml binding buffer (20 mM Tris–HCl pH 8.0, 500 mM NaCl) and then lysed with 0.5 mm diameter glass beads using a Precellys 24 homogenizer at 5000 rev min^{−1} for 6 × 30 s, cooling the mixture in an ice bath during each interval of cell lysis. After centrifugation at 15 200g for 20 min at 277 K, the clear lysate recovered as the supernatant was loaded onto an affinity-chromatography column containing 10 ml Ni–NTA agarose (GE Healthcare) resin pre-equilibrated with binding buffer. The tagged protein was eluted with 30 ml binding buffer containing 500 mM imidazole, which was then concentrated and loaded onto a HiLoad 16/600 Superdex 200 pg gel-filtration chromatography column (GE Healthcare) and eluted with 20 mM Tris–HCl pH 8.0, 200 mM NaCl (buffer A) at a flow rate of 1 ml min^{−1}. The retention volume corresponding to the target protein indicates that it is a monomer

in solution (Fig. 2*a*). The fractions corresponding to the peak were pooled and adjusted to 12 mg ml^{−1}. The purity of the target protein-containing fractions was estimated by SDS–PAGE with Coomassie staining, which revealed a single band that corresponds to the expected molecular weight (Fig. 2*b*). The protein concentration was measured with the Bradford Protein Assay Kit (Bio-Rad) using BSA as a standard and buffer A as a control. Macromolecule-production information is summarized in Table 1.

2.2. Crystallization

Preliminary screening for initial crystallization conditions was performed by the sitting-drop vapour-diffusion method using Crystal Screen, Crystal Screen 2, Index, SaltRX (Hampton Research) and ProPLEX (Molecular Dimensions) at 285 K by mixing 1 µl 12 mg ml^{−1} protein solution (in buffer A) with an equal volume of reservoir solution in 48-well plates. Multiple small needle-like crystals were obtained from several conditions. Among these conditions, the crystals grown in Crystal Screen condition No. 9 [0.2 M ammonium acetate, 0.1 M sodium citrate tribasic dihydrate pH 5.6, 30% (w/v) polyethylene glycol 4000 (PEG 4000); Fig. 3*a*] and Crystal Screen condition No. 10 [0.2 M ammonium acetate, 0.1 M sodium acetate dihydrate pH 4.6, 30% (w/v) PEG 4000; Fig. 3*b*] exhibited a relatively good appearance. However, these crystals were unsuitable for X-ray diffraction owing to their small size. Therefore, Crystal Screen condition No. 9 was optimized by varying the concentration of PEG 4000 [20–30% (w/v)] versus a pH range of 4.6–6.5 in the presence of 0.2 M ammonium acetate, and Crystal Screen condition No. 10 was optimized by varying the concentration of PEG 4000 [20–30% (w/v)] versus a pH range of 4.0–5.6 in the presence of 0.2 M ammonium acetate. Clusters of needle-like crystals of

Figure 2
(*a*) Superdex 200 gel-filtration chromatography profile of Ma-AerE. The absorbance at 280 nm is shown in blue. The peak corresponding to the target protein is indicated by a red arrow. The retention volumes corresponding to the standard protein markers conalbumin, ovalbumin, carbonic anhydrase and ribonuclease A, with molecular weights of 75, 44, 29 and 13.7 kDa, respectively, are indicated by black arrows. (*b*) SDS–PAGE analysis of purified Ma-AerE. The protein was analyzed on 10% SDS–PAGE and stained with Coomassie Blue. Lane 1, Ma-AerE pooled from the fractions of the main elution peak in gel-filtration chromatography; lane 2, molecular-weight markers labelled in kDa (Thermo Scientific).

good diffraction quality (Fig. 3c), which were used in data collection, appeared in about 5 d from 0.2 *M* ammonium acetate, 0.1 *M* sodium acetate dihydrate pH 5.6, 20%(w/v) PEG 4000.

2.3. Data collection and processing

Several pieces of the needle crystals in the cluster, which were used in X-ray diffraction experiments, were separated and mounted in CryoLoops (Hampton Research) and quickly soaked in a cryoprotectant solution consisting of 20%(v/v) glycerol, 0.2 *M* ammonium acetate, 0.1 *M* sodium acetate dihydrate pH 5.6, 20%(w/v) PEG 4000. X-ray diffraction data were collected on the BL17U beamline at the Shanghai Synchrotron Radiation Facility (SSRF) using a Quantum 315r CCD detector (Area Detector Systems Corporation). The crystals were cooled in a nitrogen stream at 100 K during data collection. The size of the beam used in the data collection was 0.1×0.1 mm. The crystal-to-detector distance was kept at 220 mm and the crystal was rotated through a total of 180° with 1° rotation per frame over 1 s. All data were indexed, integrated and scaled with *HKL-2000* (Otwinowski & Minor, 1997). The final statistics of data collection and processing from a single crystal with the best diffraction power are listed in Table 2. The diffraction data from the crystal were scaled to a maximum resolution of 1.60 Å. Analysis of the systematic absences in the diffraction data characterized the space group as $P2_12_12_1$.

2.4. Metal-element analysis

Both X-ray fluorescence (XRF) emission spectroscopy and inductively coupled plasma atomic emission spectroscopy (ICP-AES) were employed to analyze the bound metal in Ma-AerE. The XRF emission spectrum of the Ma-AerE crystal near the wavelength of the Fe *K* edge (corresponding to an energy of 7132 eV) with a scan duration of 10 s was recorded at SSRF. Purified Ma-AerE diluted to 5 mg ml⁻¹ (in buffer *A*)

Table 2

Data-collection statistics.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	0.97907
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	$a = 34.770, b = 62.133, c = 87.401$
Resolution range (Å)	50.00–1.60 (1.63–1.60)
Total No. of reflections	168245
No. of unique reflections	25640
Completeness (%)	99.4 (99.3)
Multiplicity	6.6 (6.4)
Average $I/\sigma(I)$	27.3 (3.5)
$R_{\text{merge}}^\dagger$	0.071 (0.683)
$R_{\text{p.i.m.}}^\ddagger$	0.031 (0.292)
Overall <i>B</i> factor from Wilson plot (Å ²)	19.3

$^\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, $^\ddagger R_{\text{p.i.m.}} = \frac{\sum_{hkl} [1/[N(hkl) - 1]]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the observed intensity of a reflection, $\langle I(hkl) \rangle$ is the mean intensity of reflection hkl and $N(hkl)$ is the data multiplicity.

was treated with 75%(v/v) HNO₃ for metal-element determination by ICP-AES using an Optima 2100 (PerkinElmer). An equal volume of buffer *A* was used as a control. The assay was performed two times at the Ningbo Institute of Industrial Technology (CNITECH) of the Chinese Academy of Sciences.

3. Results and discussion

Ma-AerE, a putative oxidase that participates in the biosynthetic pathway of the Choi moiety of aeruginosin, was successfully cloned and expressed in *E. coli* and purified to homogeneity by two chromatographic steps. The molecular weight of 24.51 kDa for monomeric Ma-AerE with a His₆LE tag predicted from the sequence was confirmed by 10% SDS-PAGE (Fig. 2b). The condition in which crystals of Ma-AerE with the best quality appeared was 0.2 *M* ammonium acetate, 0.1 *M* sodium acetate dihydrate pH 5.6, 20%(w/v) PEG 4000.

The XRF emission spectrum of the Ma-AerE crystal illustrates that Ag, Ca, Mn, Fe, Cd and Co are abundant in the sample (Supplementary Fig. S1). For further identification of the bound metal in Ma-AerE, ICP-AES was applied to determine the metal concentrations in an Ma-AerE solution. The results of ICP-AES show that the net concentration of Fe after subtraction of the control value is significantly higher than those of Mn and Co, and the ratio between the net molar concentration of Fe and the molar concentration of Ma-AerE is about 0.7 (Supplementary Table S1). Taken together, these results suggest that Fe is the probable bound metal and that only one metal ion is coordinated by the Ma-AerE molecule.

The diffraction data from the crystal of Ma-AerE were scaled to a maximum resolution of 1.60 Å. The crystal belonged to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters $a = 34.770, b = 62.133, c = 87.401$ Å. The resulting overall R_{merge} and $R_{\text{p.i.m.}}$ values are 7.1 and 3.1%, respectively. The mosaicity of the crystal ranges from 0.613 to 0.848° . The calculated Matthews coefficient and solvent content (Matthews, 1968) were $1.93 \text{ Å}^3 \text{ Da}^{-1}$ and 36.30%, respectively, assuming the presence of one molecule per crystal asymmetric unit. The initial structural model was

Figure 3

Crystals of Ma-AerE observed from Crystal Screen condition Nos. 9 (*a*) and 10 (*b*). (*c*) Crystals of Ma-AerE observed from the optimized condition 0.2 *M* ammonium acetate, 0.1 *M* sodium acetate dihydrate pH 5.6, 20%(w/v) PEG 4000, which were used in data collection.

obtained by molecular replacement using the coordinates of the monomer from the crystal structure of BacB from *B. subtilis* (PDB entry 3h7j; Rajavel *et al.*, 2009; 29% sequence identity, 49% sequence similarity) as the search template in *Phaser* (McCoy *et al.*, 2007) in the resolution range 50.00–3.00 Å. The structural model from the solution from *Phaser* with the highest *Z*-score and log-likelihood gain values (RFZ = 5.5, TFZ = 8.7, LLG = 62) was then refined using restrained refinement in *REFMAC5* (Murshudov *et al.*, 2011), which decreased the *R* and *R*_{free} factors to 46.4 and 51.6%, respectively, with a resulting figure of merit (FOM) of 0.398. For the convenience of subsequent refinement, two rounds of iterative molecular replacement using the refined structure as the search template in *Phaser* alternating with restrained refinement in *REFMAC5* were carried out. After restrained refinement of the model obtained from the last molecular-replacement iteration, the *R* and *R*_{free} factors had decreased to 38.0 and 42.2%, respectively, with a resulting FOM of 0.605. At this stage, the Ma-AerE molecule was well fitted by the electron density, allowing manual refinement to be conveniently performed using *Coot* (Emsley & Cowtan, 2004). The final model building and characterization of the catalytic specificity of Ma-AerE are ongoing.

Acknowledgements

We thank the staff of the BL17U beamline at SSRF for their assistance in data collection. Financial support for this work was provided by the National Natural Science Foundation of China (Grant No. 31400683), the Natural Science Foundation of Zhejiang Province (Grant No. LQ14C050001), the Subject Project of Ningbo University (Grant No. xkl1343) and the Talent Plan of Ningbo University (Grant Nos.

013-E00843134702 and 013-E00843144702). This work was also sponsored by the K. C. Wong Magna Fund of Ningbo University.

References

- Anantharaman, V., Aravind, L. & Koonin, E. V. (2003). *Curr. Opin. Chem. Biol.* **7**, 12–20.
- Dunwell, J. M., Culham, A., Carter, C. E., Sosa-Aguirre, C. R. & Goodenough, P. W. (2001). *Trends Biochem. Sci.* **26**, 740–746.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst. D* **60**, 2126–2132.
- Gopal, B., Madan, L. L., Betz, S. F. & Kossiakoff, A. A. (2005). *Biochemistry*, **44**, 193–201.
- Ishida, K., Christiansen, G., Yoshida, W. Y., Kurmayer, R., Welker, M., Valls, N., Bonjoch, J., Hertweck, C., Börner, T., Hemscheidt, T. & Dittmann, E. (2007). *Chem. Biol.* **14**, 565–576.
- Ishida, K., Welker, M., Christiansen, G., Cadel-Six, S., Bouchier, C., Dittmann, E., Hertweck, C. & Tandeau de Marsac, N. (2009). *Appl. Environ. Microbiol.* **75**, 2017–2026.
- Mahlstedt, S., Fielding, E. N., Moore, B. S. & Walsh, C. T. (2010). *Biochemistry*, **49**, 9021–9023.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J. (2007). *Acta Cryst. D* **63**, 32–41.
- 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.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Rajavel, M., Mitra, A. & Gopal, B. (2009). *J. Biol. Chem.* **284**, 31882–31892.
- Rios Steiner, J. L., Murakami, M. & Tulinsky, A. (1998). *J. Am. Chem. Soc.* **120**, 597–598.
- Sandler, B., Murakami, M. & Clardy, J. (1998). *J. Am. Chem. Soc.* **120**, 595–596.
- Welker, M., Maršálek, B., Šejnohová, L. & von Döhren, H. (2006). *Peptides*, **27**, 2090–2103.
- Welker, M. & von Döhren, H. (2006). *FEMS Microbiol. Rev.* **30**, 530–563.