Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of AerF from *Microcystis aeruginosa*, a putative reductase participating in aeruginosin biosynthesis

Ruyi Ding,^{a,b,†} Cui Xu,^{a,b,†} Xu Chen,^a Mengyun Bao^{a,b} and Xiaoting Qiu^{a,*}

Received 11 February 2015

Accepted 12 March 2015

Edited by N. Sträter, University of Leipzig, Germany

† These authors contributed equally to this work.

Keywords: *Microcystis aeruginosa*; aeruginosin; 2-carboxy-6-hydroxyoctahydroindole moiety; AerF.

^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

The 2-carboxy-6-hydroxyoctahydroindole moiety is an essential residue for the antithrombotic activity of aeruginosins, which are a class of cyanobacteria-derived bioactive linear tetrapeptides. The biosynthetic pathway of the 2-carboxy-6-hydroxyoctahydroindole moiety has not yet been resolved. AerF was indicated to be involved in the biosynthesis of the 2-carboxy-6-hydroxyoctahydroindole moiety. This study reports the cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of AerF from *Microcystis aeruginosa* with a C-terminal His₆ tag. The crystal diffracted to a maximum resolution of 1.38 Å and belonged to the tetragonal space group *P*₄₃₂₂, with unit-cell parameters *a* = *b* = 101.581, *c* = 116.094 Å. The calculated Matthews coefficient and solvent content of the crystal were 2.47 Å³ Da^{−1} and 50.32%, respectively. The initial model of the structure was obtained by the molecular-replacement method and refinement of the structure is in progress.

1. Introduction

A number of natural products have been isolated from cyanobacteria. Investigations of the biosynthetic pathways of these compounds have revealed some fascinating mechanisms that have never been found in other microorganisms (Kehr *et al.*, 2011). Among these cyanobacteria-derived natural products, bioactive oligopeptides, including both linear and cyclic peptides, have been the focus of research (Welker & von Döhren, 2006). Aeruginosins constitute a class of highly variable linear tetrapeptides, which have been isolated from freshwater cyanobacterial genera and marine sponges (Ishida *et al.*, 2007). Aeruginosins are specific inhibitors of different types of serine-type proteases (Sandler *et al.*, 1998; Rios Steiner *et al.*, 1998); therefore, they are considered to be potential antithrombotic agents (Wang & Goyal, 2009). The unique physiological properties of aeruginosin are mainly contributed by the 2-carboxy-6-hydroxyoctahydroindole (Choi) moiety. Choi is a bicyclic amino acid; it is the hallmark of aeruginosin and is present in all variants of aeruginosins (Ishida *et al.*, 2009). When aeruginosin binds to a serine-type protease, the Choi moiety directly interacts with the active site of the protease; thus it is the core moiety that determines the inhibition efficiency of aeruginosin (Wang & Goyal, 2009).

Three genes corresponding to Choi biosynthesis, *aerD*, *aerE* and *aerF*, have been characterized in the freshwater cyanobacteria *Planktothrix agardhii* and *Microcystis aeruginosa* via sequence alignment and genetic rescue experiments (Ishida *et al.*, 2007). AerD, the product of *aerD*, has been illustrated to

© 2015 International Union of Crystallography

Table 1
Macromolecule-production information.

Source organism and DNA source	<i>M. aeruginosa</i> strain PCC7806
Forward primer†	GGAATTCATATGGATTAGGATTAAGATAAAA-GTGG
Reverse primer‡	CCGCTCGAGAAATGCTGACGAGGATAAACCC
Cloning vector and expression vector	pET-22b
Expression host	<i>E. coli</i> BL21 (DE3)
Complete amino-acid sequence of the construct produced§	MDLGLKDKVALITGSSAGIGFTIAEKLAEEGCHL-IICGRNSQRLEQAYQSLAQAYPAQQLRLVAD-VHQAQDSEQLIQDSLNGYKIDILVNNSEGAN-FAENLIENLSDEDWLVNFQKGLIGYIRLTNLV-LPIMKKQHWGRIVNIIGTSGKEPSRLVKSGV-VNAALMNFTKSVARQTAPYVNLINSVNPQVID-TPRHREYLEIYAKKEGTTPLIRERILKTIPTM-NRIGTTEEFANLVVFLASECASYITGITIPLD-GGLSSSAFLHHHHH

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

be an enzyme that catalyzes the non-aromatizing decarboxylation of prephenate, which was considered to catalyze the first step in the pathway from prephenate to Choi (Mahlstedt *et al.*, 2010), while the exact roles of AerE and AerF, the products of *aerE* and *aerF*, in Choi biosynthesis remain to be elucidated. Based on sequence similarity, AerF belongs to the short-chain dehydrogenase/reductase (SDR) family of enzymes. Enzymes of the SDR family have a common α/β Rossmann fold for NAD(P)H binding and catalyze different substrates through similar mechanisms (Kavanagh *et al.*, 2008). Several enzymes of the SDR family have been structurally and functionally elucidated, such as YwfH from *Bacillus subtilis* (Rajavel *et al.*, 2013), L-sorbose reductase (SR) from *Gluconobacter frateurii* (Kubota *et al.*, 2011), SDR (APE0912) from *Aeropyrum pernix* (Yamamura *et al.*, 2008)

and 3-oxoacyl-(acyl-carrier-protein) reductase (FabG) from *Pseudomonas aeruginosa* (Cukier *et al.*, 2013). Sequence alignment with these enzymes illustrates that most of the residues participating in NAD(P)H binding are conserved in AerF, while some critical catalytic residues are not conserved in AerF (Fig. 1). It suggests that AerF is probably an NAD(P)H-dependent reductase with a different substrate specificity compared with other members of the SDR family, which should be revealed by structural information on AerF. Towards this purpose, AerF from *Microcystis aeruginosa* (Ma-AerF) was cloned and expressed in *Escherichia coli*, purified *via* two steps and crystallized; an X-ray diffraction data set was collected from the crystals, thus preparing for further study of the substrate specificity and catalytic mechanism of AerF and providing a foundation for exploring the role of AerF in Choi biosynthesis.

2. Materials and methods

2.1. Cloning, expression and purification

The open reading frame of full-length Ma-AerF was obtained *via* PCR from the genome of *M. aeruginosa* (strain PCC7806). It was cloned into pET-22b vector (Novagen), digested using *Nde*I and *Xho*I (New England Biolabs) to create recombinant Ma-AerF with a C-terminal hexahistidine tag (LEHHHHHH; Table 1). After the sequence of the open reading frame of Ma-AerF had been verified by DNA sequencing (Invitrogen), the vector was transformed into *E. coli* BL21 (DE3) cells (Novagen). The bacteria harbouring the expression vector were cultured in 16 ml LB medium overnight and then used to inoculate 1.6 l medium containing 50 $\mu\text{g ml}^{-1}$ ampicillin. The cells were grown at 310 K for 2.5 h until the OD_{600 nm} reached 0.8 and protein expression was

Figure 1

Multiple sequence alignment of Ma-AerF with structurally and functionally characterized enzymes of the SDR family mentioned in §1. The NAD(P)H-binding residues and critical catalytic residues characterized in other SDR enzymes are indicated with green and red stars, respectively. The C-terminal sequences were not included in the alignment. The alignment was generated using the *ClustalW* web server and this figure was prepared using *EScript*.

then induced for 24 h with 0.25 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) at 289 K. The cell pellet was collected by centrifugation at 9990g for 5 min and resuspended in 30 ml binding buffer (20 mM Tris–HCl pH 8.0, 500 mM NaCl); it was then disrupted with 0.5 mm-diameter glass beads using a Fast-prep (Precellys 24) at a speed of 5000 rev min⁻¹ for 6 $\times$ 30 s. An ice bath was used to cool the sample after each interval of cell lysis. After centrifugation at 15 200g for 20 min, the supernatant recovered from the lysate was loaded onto 10 ml Ni–NTA agarose (GE Healthcare) resin which had been pre-equilibrated with binding buffer. The tagged protein was eluted with 30 ml binding buffer containing 500 mM imidazole; it was then concentrated for further purification using a

HiLoad 16/600 Superdex 200 pg gel-filtration column (GE Healthcare), from which it was eluted with 20 mM Tris–HCl pH 8.0, 200 mM NaCl (buffer A) at a flow rate of 1 ml min⁻¹. The gel-filtration column was calibrated using thyroglobulin, ferritin, aldolase, conalbumin, ovalbumin, carbonic anhydrase, ribonuclease A and aprotinin, with molecular weights of 669, 440, 158, 75, 44, 29, 13.7 and 6.5 kDa, respectively. Blue dextran 2000 was used to determine the void volume of the column. There is a peak located immediately near the main peak and thus only the fractions of the main peak that were distant from this peak were pooled (as indicated in Fig. 2*a*) and concentrated to 25 mg ml⁻¹ using a 10 kDa cutoff Amicon centrifugal ultrafilter concentrator (Millipore). Examination of the purified protein by SDS–PAGE revealed a single band that corresponded to the expected molecular weight (Fig. 2*b*). Protein concentration was measured with the Bradford Protein Assay Kit (Bio-Rad).

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. 1 μ l protein solution (25 mg ml⁻¹ in buffer A) was mixed with an equal volume of reservoir solution in 48-well plates and the drops were equilibrated against 100 μ l reservoir solution. Crystals appeared from several conditions. Among these conditions, the crystals obtained using Crystal Screen condition No. 14 [0.2 M calcium chloride, 0.1 M sodium HEPES pH 7.5, 28%(v/v) polyethylene glycol 400] had the best appearance. This condition was further optimized using various concentrations of PEG 400 [15–30%(v/v)] *versus* a pH range of 7.0–8.0 in the presence or absence of 0.2 M calcium chloride. Crystals of the best diffraction quality appeared in about 2 d from the condition 0.2 M calcium chloride, 0.1 M

Figure 2

(*a*) Gel-filtration chromatography profile of Ma-AerF. Absorbance at 280 nm is shown in blue. The *x* axis starts from the elution volume corresponding to sampling (26.96 ml). The range of fractions pooled for concentration is indicated by red dashed lines. (*b*) SDS–PAGE of purified Ma-AerF. The protein was analyzed on 10% SDS–PAGE and stained with Coomassie Blue. Lane 1, molecular-weight markers (labelled in kDa; Thermo Scientific); lane 2, Ma-AerF obtained from two purification steps.

Figure 3

Crystals of Ma-AerF, as grown in 0.2 M calcium chloride, 0.1 M sodium HEPES pH 7.0, 23%(v/v) polyethylene glycol 400, which were used in the diffraction experiment.

Table 2

Data collection and processing.

Values in parentheses are for the outer shell.

Wavelength (Å)	0.97907
Temperature (K)	100
Crystal-to-detector distance (mm)	190
Space group	$P4_322$
Unit-cell parameters (Å)	$a = b = 101.581$, $c = 116.094$
Resolution range (Å)	45.00–1.38 (1.40–1.38)
Total No. of reflections	1678642
No. of unique reflections	122470 (5628)
Completeness (%)	98.5 (92.1)
Multiplicity	13.7 (9.4)
Average $I/\sigma(I)$	33.9 (3.3)
$R_{\text{merge}}^\dagger$	0.083 (0.906)
$R_{\text{p.i.m.}}^\ddagger$	0.023 (0.301)
Overall B factor from Wilson plot (Å ²)	15.0

$^\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, $^\ddagger R_{\text{p.i.m.}} = \frac{\sum_{hkl} [1/N(hkl) - 1]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, 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 redundancy.

sodium HEPES pH 7.0, 23%(v/v) polyethylene glycol 400 (Fig. 3). These crystals were used for data collection.

2.3. Data collection and processing

Before data collection, the crystals were soaked for a short time in a cryoprotectant solution consisting of 20%(v/v) glycerol, 0.2 *M* calcium chloride, 0.1 *M* sodium HEPES pH 7.0, 23%(v/v) polyethylene glycol 400 and flash-cooled in a nitrogen stream at 100 K. X-ray diffraction data were collected on beamline 17U1 of the Shanghai Synchrotron Radiation Facility (SSRF) using a Quantum 315r CCD detector (Area Detector Systems Corporation). The crystal-to-detector distance was kept at 190 mm and the crystal was rotated through a total of 180° with 0.5° rotation per frame over 0.5 s. All data were indexed, integrated and scaled with *HKL-2000* (Otwinowski & Minor, 1997). Analysis of the systematic absences in the diffraction data narrowed the choice of possible space groups to $P4_222$, $P4_122$ or $P4_322$. Inspection of the molecular-replacement solution obtained using *Phaser* (McCoy *et al.*, 2007) as implemented in the *CCP4* suite (Winn *et al.*, 2011) confirmed the space group as $P4_322$. The final statistics of data collection and processing are given in Table 2.

3. Results and discussion

AerF from *M. aeruginosa*, which is a putative reductase that participates in the biosynthetic pathway of the Choi moiety of aeruginosin, was successfully cloned and expressed in *E. coli* and purified to homogeneity with a C-terminal His₆ tag. The molecular weight of monomeric His₆ Ma-AerF of 30.3 kDa predicted from the sequence was confirmed by 10% SDS-PAGE (Fig. 2*b*). The retention volume of the main peak in the gel-filtration chromatography profile corresponding to the target protein indicates that it is a tetramer in solution (theoretical molecular weight of tetrameric His₆-tagged Ma-AerF is 121.2 kDa and the apparent molecular weight calcu-

lated from the retention volume is 107.4 kDa; Fig. 2*a*). The crystals of Ma-AerF with best diffraction power were crystallized from 0.2 *M* calcium chloride, 0.1 *M* HEPES sodium pH 7.0, 23%(v/v) polyethylene glycol 400 (Fig. 3).

The crystal of Ma-AerF diffracted to a maximum resolution of 1.38 Å and belonged to the tetragonal space group $P4_322$, with unit-cell parameters $a = b = 101.581$, $c = 116.094$ Å (Table 2). The resulting overall R_{merge} and $R_{\text{p.i.m.}}$ are 8.3 and 2.3%, respectively. The mosaicity of the crystal ranged from 0.264 to 0.442°. Cell-content analysis indicated that there are two molecules per asymmetric unit, with a solvent content of 50.32% and a Matthews coefficient (Matthews, 1968) of 2.47 Å³ Da⁻¹. An initial model for the structure was obtained by molecular replacement using the coordinates of the monomer of the crystal structure of YwfH from *Bacillus subtilis* (PDB entry 3u49; Rajavel *et al.*, 2013; 36% sequence identity) as the input model and a search for two monomers in the asymmetric unit via *Phaser* (McCoy *et al.*, 2007) in the resolution range 45.00–3.50 Å. Both monomers were found in the asymmetric unit, which is consistent with the result of the Matthews-coefficient calculation. The correctness of the solution of *Phaser* with space group $P4_322$ was clearly indicated by highest Z-scores and log-likelihood gain values for one monomer (RFZ = 4.9, TFZ = 8.3, LLG = 45) and the other monomer (RFZ = 3.2, TFZ = 8.4, LLG = 146). The model was then refined using restrained refinement by *REFMAC5* (Murshudov *et al.*, 2011), which decreased the R factor and R_{free} factor to 37.7 and 41.1%, respectively, with a resulting figure of merit of 0.63. The electron density was clear for the majority of the molecules, allowing subsequent refinement to be conveniently conducted. After several rounds of refinement in the resolution range 45.00–1.38 Å using restrained refinement by *REFMAC5* and manual refinement by *Coot* (Emsley & Cowtan, 2004), both the resulting R factor and R_{free} factor decreased to below 25%. Final model building is currently in progress.

Acknowledgements

We thank the staff of beamline 17U1 at the 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), a Subject Project of Ningbo University (grant No. xkl1343) and the Talent Plan of Ningbo University. This work was also sponsored by the K. C. Wong Magna Fund at Ningbo University.

References

- Cukier, C. D., Hope, A., Elamin, A., Moynie, L., Schnell, R., Schach, S., Kneuper, H., Singh, M., Naismith, J., Lindqvist, Y., Gray, D. & Schneider, G. (2013). *Chem. Biol.* **8**, 2518–2527.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- 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.

- Kavanagh, K., Jörnvall, H., Persson, B. & Oppermann, U. (2008). *Cell. Mol. Life Sci.* **65**, 3895–3906.
- Kehr, J., Gatte Picchi, D. & Dittmann, E. (2011). *Beilstein J. Org. Chem.* **7**, 1622–1635.
- Kubota, K., Nagata, K., Okai, M., Miyazono, K., Soemphol, W., Ohtsuka, J., Yamamura, A., Saichana, N., Toyama, H., Matsushita, K. & Tanokura, M. (2011). *J. Mol. Biol.* **407**, 543–555.
- 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., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- 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., Perinbam, K. & Gopal, B. (2013). *Acta Cryst. D* **69**, 324–332.
- 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.
- Wang, G. & Goyal, N. (2009). *Cardiovasc. Hematol. Agents Med. Chem.* **7**, 147–165.
- Welker, M. & von Döhren, H. (2006). *FEMS Microbiol. Rev.* **30**, 530–563.
- Winn, M. D. *et al.* (2011). *Acta Cryst. D* **50**, 760–763.
- Yamamura, A., Ichimura, T., Mimoto, F., Ohtsuka, J., Miyazono, K., Okai, M., Kamo, M., Lee, W., Nagata, K. & Tanokura, M. (2008). *Proteins*, **70**, 1640–1645.