Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

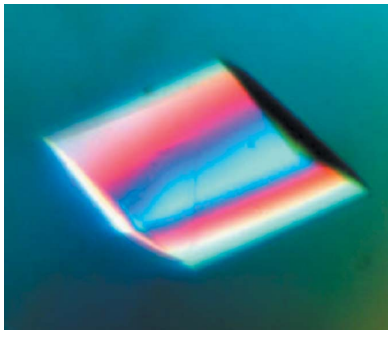
A putative polyketide-synthesis protein XC5357 from *Xanthomonas campestris*: heterologous expression, crystallization and preliminary X-ray analysis

Chiao-Li Chu,[a] Ko-Hsin Chin,[a] Fu-Yang Lin,[a] Chia-Cheng Chou,[b,c] Cheng-Chung Lee,[b,c] Hui-Lin Shr,[b,c] Ping-Chiang Lyu,[d] Andrew H.-J. Wang[b,c] and Shan-Ho Chou[a]*

[a]Institute of Biochemistry, National Chung-Hsing University, Taichung 40227, Taiwan, [b]Institute of Biological Chemistry, Academia Sinica, Nankang, Taipei, Taiwan, [c]Core Facility for Protein Crystallography, Academia Sinica, Nankang, Taipei, Taiwan, and [d]Department of Life Science, National Tsing Hua University, Hsin-Chu, Taiwan

Correspondence e-mail: shchou@nchu.edu.tw

Received 8 April 2005
Accepted 14 June 2005
Online 30 June 2005



Xanthomonas campestris pv. *campestris* (Xcc) is a Gram-negative yellow-pigmented bacterium and is the causative agent of black rot, one of the major worldwide diseases of cruciferous crops. It also synthesizes a variety of polyketide metabolites that lead to important antibiotics. XC5357 is a putative 12.2 kDa protein of unknown structure from Xcc that is likely to be essential for polyketide synthesis. It was overexpressed in *Escherichia coli*, purified and crystallized using the hanging-drop vapour-diffusion method. The crystals belong to the triclinic space group $P1$, with unit-cell parameters $a = 43.7$, $b = 43.7$, $c = 46.5$ Å, $\alpha = 65.0$, $\beta = 64.9$, $\gamma = 73.4^\circ$, and diffracted to a resolution of 1.85 Å.

1. Introduction

Microorganisms and plants synthesize a large variety of polyketide metabolites, many of which are medically important antibiotics or exhibit other pharmacological properties such as antibacterial, antiviral or antitumour properties (Hopwood, 1997). Although the biosyntheses of such polyketide products involve many steps, methods have been developed to reconstitute active protein complexes to produce useful antibiotics (Bao *et al.*, 1998). There is thus a broad interest in trying to understand their detailed biosynthetic pathways in the hope of producing novel polyketide-derived drugs using genetically engineered polyketide synthases (PKSs) produced *via* recombinant microorganisms. To achieve this goal, it is crucial to be able to investigate their biosynthesis mechanism from both the genetic and structural points of view (Cane *et al.*, 1998; Shen, 2000).

Tetracenomycin (TCM) C is a cytotoxic antibiotic produced by *Streptomyces glaucescens* and is notable for its broad activity against actinomycetes (Hutchinson, 1997). Its synthesis is currently one of the models for antibiotic production in *Streptomyces* species. Sequence analysis of the entire TCM C gene-cluster *tcmGHIJKLMNO* operon has been completed and a model for Tcm PKS-catalyzed synthesis of TCM C has also been created (Shen & Hutchinson, 1993). A reconstitution experiment of the PKS for TCM F2, a precursor of TCM C, suggests that the active PKS complex consists of at least four major proteins, including the Tcm*K*, Tcm*L*, Tcm*M* and Tcm*N* gene products (Bao *et al.*, 1998). Tcm*J* is also one of the components of the PKS complex, but its function is currently unknown, although its addition to the Tcm*KLMN* complex can increase the production of TCM F2 by nearly fourfold (Shen & Hutchinson, 1993).

XC5357 (gi|21112129) from a local plant pathogen *Xanthomonas campestris* pv. *campestris* strain 17 has been annotated as a polyketide-synthesis protein using a bioinformatics approach (http://xcc.life.nthu.edu.tw/; da Silva *et al.*, 2002). It consists of 113 amino acids and shares 100% identity with a similar protein from *X. campestris* pv. *campestris* ATCC33913 (gi|21112129; da Silva *et al.*, 2002), 34% identity with a protein from *Mesorhizobium loti* MAFF303099 (Kaneko *et al.*, 2000) and 32% identity (55% similarity) with the Tcm*J* protein from *S. glaucescens* (Bentley *et al.*, 2002; Salanoubat *et al.*, 2002). To date, no tertiary structure for the Tcm*J*-like proteins has been reported. In this manuscript, we describe the cloning, purification, crystallization and initial X-ray analysis of XC5357.

2. Materials and methods

2.1. Cloning, expression and purification

The XC5357 gene fragment was amplified using sticky PCR (Zeng, 1998) from a local Xcc strain. Two PCR reactions, one with the forward primer P1-**AATTCG**GAATGCAGTACGCAACGTTG-GAATTG) and the backward primer P2-**G**GCCTTCGCCTGCCG-GCAATGGTTC, and another with the forward primer P3-**C**GGAATGCAGTACGCAACGTTGGAATTGAAC and the back-ward primer P4-**TCGAG**GCCTTCGCCTGCCGGCAATG, were carried out separately. After amplification, the PCR products were combined and denatured at 368 K for 5 min before cooling to obtain approximately 25% of the correct PCR products containing hanging *Eco*RI 5′AATTC/G-5′ and *Xho*I 5′-TCGAG/C-5′ cohesive ends. They were then ligated with a modified pET-32a(+) vector (Shih *et al.*, 2002) cut with the same enzymes and cloned into the *Escherichia coli* BL21 (DE3) host cell. The final construct codes for a thioredoxin tag protein (109 amino acids), an XC5357 protein (113 amino acids) and a C-terminal His_6 tag under the control of a T7 promoter. The transformed *E. coli* BL21 (DE3) host cell was grown in LB medium at 310 K until an OD of 0.8 was attained. Overexpression of the fusion protein was induced by the addition of 0.5 m*M* IPTG at 293 K for 20 h. The cells were harvested, resuspended in equilibration buffer (20 m*M* Na_2HPO_4, 70 m*M* NaCl pH 8.5) and lysed using a micro-fluidizer (Microfluidics). After centrifugation, the tagged fusion protein was purified by immobilized metal-affinity chromatography (IMAC) on a cobalt column (BD Biosciences). The fusion protein was eluted with 20 m*M* Tris pH 8.0, 70 m*M* NaCl and a gradient of 100–300 m*M* imidazole. The fractions containing XC5357 were dialyzed repeatedly against 140 m*M* NaCl, 2.7 m*M* KCl, 10 m*M* Na_2HPO_4 and 1.8 m*M* KH_2PO_4 pH 8.0. After concentration, the mixture was loaded onto a cobalt column and the thioredoxin tag was removed from the bound fusion protein by thrombin cleavage at 295 K for 16 h. The XC5357 target protein was then eluted with 20 m*M* Tris pH 8.0, 70 m*M* NaCl and a gradient of 100–300 m*M* imidazole. The desired fractions were dialyzed repeatedly in 20 m*M* Tris pH 8.0 and 70 m*M* NaCl. For crystallization, the XC5357 protein was further purified using an anion-exchange column (Pharmacia Inc.). The fractions eluted with 20 m*M* Tris pH 8.0, 550 m*M* NaCl were combined and dialyzed against 20 m*M* Tris pH 8.0 and 70 m*M* NaCl. The final construct (130 amino acids) contains the target protein (113 amino acids), an extra nonapeptide (GSGGGGEFG) at the N-terminal end and an extra octapeptide LEH_6 at the C-terminal end with an expected molecular weight of 13 937 Da, which was confirmed by mass-spectrometric analysis (13 935 Da). The over-expression and purification of XC5357 was monitored on SDS–PAGE as shown in Fig. 1.

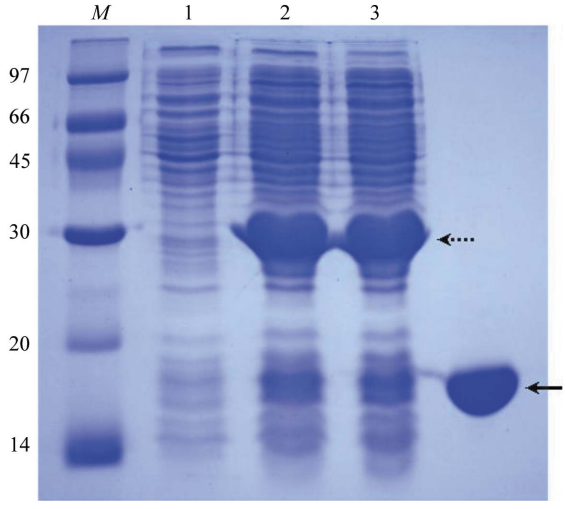


Figure 1
SDS–PAGE showing the overexpression and purification of XC5357. Lane *M*, molecular-weight marker standards in kDa; lane 1, whole cell lysate before IPTG induction; lane 2, whole cell lysate after IPTG induction; lane 3, soluble fraction after IPTG induction; lane 4, His-tag column-purified XC5357 after thrombin cleavage. The dotted arrow indicates the position of Trx-fused XC5357 and the solid arrow the position of the Trx-cleaved XC5357.

Table 1
Data-collection statistics for XC5357.

Values in parentheses are for the highest resolution shell.

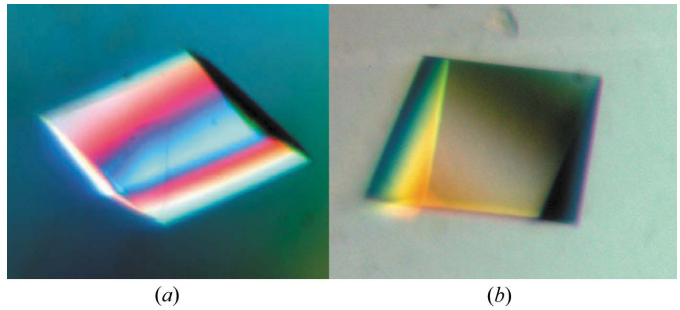
Space group	*P*1
Unit-cell parameters (Å, °)	a = 43.7, b = 43.7, c = 46.5, α = 65.0, β = 64.9, γ = 73.4
Temperature (K)	100
Wavelength (Å)	1.5418
Resolution range (Å)	24.0–1.85 (1.92–1.85)
Mosaicity (°)	0.6
Unique reflections	22322 (2144)
Redundancy	1.7 (1.7)
Completeness (%)	93.4 (90.3)
R_{merge} (%)	2.9 (5.0)
Mean $I/\sigma(I)$	23.3 (17.3)
Solvent content (%)	58.1 or 37.1

2.2. Crystallization

For crystallization, the protein was concentrated to 25 mg ml^{-1} in 20 m*M* Tris pH 8.0 and 70 m*M* NaCl using an Amicon Ultra-10 (Millipore). Screening for crystallization conditions was performed using sitting-drop vapour diffusion in 96-well plates (Hampton Research) at 295 K by mixing 0.5 µl protein solution with 0.5 µl reagent solution. Initial screens included the Hampton sparse-matrix Crystal Screens 1 and 2, a systematic PEG–pH screen and the PEG/Ion Screen and were performed using a Gilson C240 crystallization workstation. Prism-shaped crystals appeared in 3 d from a reservoir solution comprising 0.1 *M* citric acid buffer pH 5.0 and 2.4 *M* ammonium sulfate. This initial conditions were then optimized by varying the concentration of ammonium sulfate. Crystals suitable for diffraction experiments were grown by mixing 1.5 µl protein solution with 1.5 µl reagent solution and reached maximum dimensions of 0.3 × 0.3 × 0.1 mm after one week (Fig. 2).

2.3. Data collection

Crystals were soaked in a cryoprotectant solution comprising reservoir solution plus 25%(*v*/*v*) glycerol. X-ray diffraction data were



(*a*) (*b*)

Figure 2
Crystallization of XC5357 from *X. campestris* by the hanging-drop vapour-diffusion method. The final optimized crystallization conditions were 0.1 *M* citric acid buffer pH 5.0, 2.2 *M* $(NH_4)_2SO_4$ (*a*). Crystals could also be grown in reagent solution containing only 2.2 *M* $(NH_4)_2SO_4$ (*b*). The approximate dimensions of these crystals were 0.3 × 0.3 × 0.1 mm.

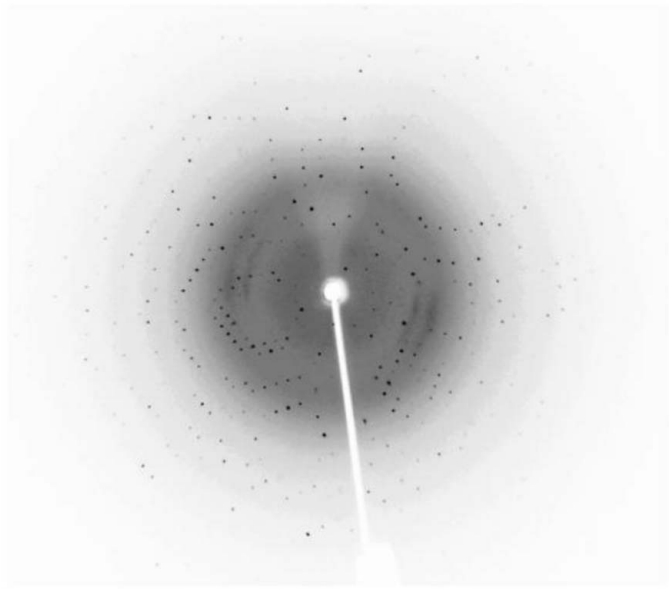


Figure 3
Diffraction pattern collected in-house from a flash-frozen XC5357 crystal with 25% glycerol cryoprotectant. The exposure time was 10 min, with an oscillation range of 1.0° and a crystal-to-detector distance of 90 mm.

collected using Cu $K\alpha$ radiation from a Rigaku MicroMax007 rotating-anode generator equipped with Osmic mirror optics and an R-AXIS IV^{++} image plate. A native data set was collected to a maximum resolution of 1.85 Å. The data were indexed and integrated using the *HKL* software suite (Otwinowski & Minor, 1997), giving a data set of 93.4% completeness with an overall $R_{\rm merge}$ of 2.9% on intensities. The crystals belong to the triclinic space group $P1$, with either two or three molecules in the asymmetric unit, corresponding to 58.1 or 37.1% solvent content, respectively. The data-collection statistics are summarized in Table 1. An X-ray diffraction image collected in-house is shown in Fig. 3.

3. Results and discussion

The gene sequence of XC5357 was confirmed after cloning and consists of 342 bp coding for 113 amino-acid residues. The purified XC5357 contains an extra nonapeptide (GSGGGGEFG) at the N-terminal end and an extra octapeptide LEH_6 at the C-terminal end and shows a greater than 99% purity, with a single band of approximately 15 kDa on SDS–PAGE (Fig. 1). The extra 17 amino acids at the ends did not seem to affect the crystallization process, as crystals readily grew overnight to dimensions of 0.3 × 0.3 × 0.1 mm (Fig. 2). The crystals diffracted to a resolution of 1.85 Å (Fig. 3). A gel-filtration experiment indicated a molecular weight of approximately 35 kDa, suggesting that XC5357 forms a dimer in solution.

The high-resolution diffraction obtained from the native crystals establishes their suitability for X-ray structural analysis. To date, there is no tertiary structural information available for the Tcm*J*-like proteins. We now plan to solve the structure of XC5357 by the multiwavelength anomalous diffraction (MAD) method (Hendrickson & Ogata, 1997; Terwilliger & Berendzen, 1999) using selenomethionine-substituted protein, given that XC5357 contains two methionines.

This work was supported by an Academic Excellence Pursuit grant from the Ministry of Education and by the National Science Council, Taiwan to S-HC and P-CL. We also thank the Core Facilities for Protein Production of the Academia Sinica, Taiwan for providing us with the original vectors used in this study and the Core Facilities for Protein X-ray Crystallography of the Academia Sinica, Taiwan for assistance in preliminary X-ray analysis.

References

Bao, W., Wendt-Pienkowski, E. & Hutchinson, C. R. (1998). *Biochemistry*, **37**, 8132–8138.
Bentley, S. D. *et al.* (2002). *Nature (London)*, **417**, 141–147.
Cane, D. E., Walsh, C. T. & Khosla, C. (1998). *Science*, **282**, 63–68.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
Hopwood, D. A. (1997). *Chem. Rev.* **97**, 2465–2497.
Hutchinson, C. R. (1997). *Chem. Rev.* **97**, 2525–2535.
Kaneko, T. *et al.* (2000). *DNA Res.* **7**, 331–338.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Salanoubat, M. *et al.* (2002). *Nature (London)*, **415**, 497–502.
Shen, B. (2000). *Curr. Opin. Chem. Biol.* **7**, 285–295.
Shen, B. & Hutchinson, C. R. (1993). *Science*, **262**, 1535–1540.
Shih, Y.-P., Kung, W.-M., Chen, J.-C., Yeh, C.-H., Wang, A. H.-J. & Wang, T.-F. (2002). *Protein Sci.* **11**, 1714–1719.
Silva, A. C. R. da *et al.* (2002). *Nature (London)*, **417**, 459–463.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Zeng, G. (1998). *Biotechniques*, **25**, 206–208.