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

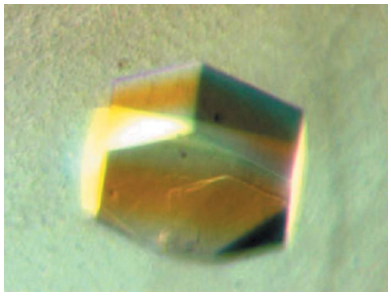
Preparation, crystallization and preliminary X-ray characterization of a conserved hypothetical protein XC1692 from *Xanthomonas campestris*

Ko-Hsin Chin,[a] Zhao-Wei Huang,[a] Kun-Chou Wei,[a] Chia-Cheng Chou,[b,c] Cheng-Chung Lee,[b,c] Hui-Lin Shr,[b,c] Fei Philip Gao,[d] Ping-Chiang Lyu,[e] 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, [d]National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL 32310, USA, and [e]Department of Life Science, National Tsing Hua University, Hsin-Chu, Taiwan

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

Received 8 April 2005
Accepted 13 June 2005
Online 30 June 2005



Xanthomonas campestris pv. *campestris* strain 17 is a Gram-negative yellow-pigmented pathogenic bacterium that causes black rot, one of the major worldwide diseases of cruciferous crops. Its genome contains approximately 4500 genes, one third of which have no known structure and/or function yet are highly conserved among several different bacterial genuses. One of these gene products is XC1692 protein, containing 141 amino acids. It was overexpressed in *Escherichia coli*, purified and crystallized in a variety of forms using the hanging-drop vapour-diffusion method. The crystals diffract to at least 1.45 Å resolution. They are hexagonal and belong to space group $P6_3$, with unit-cell parameters $a = b = 56.9$, $c = 71.0$ Å. They contain one molecule per asymmetric unit.

1. Introduction

One of the major goals of a structural genomics program is to provide an abundance of three-dimensional protein structures in order to allow better understanding of the protein sequence–structure–function relationship (Zarembinski *et al.*, 1998; Shin *et al.*, 2002; Pal & Eisenberg, 2005). From structural information on its fold, motifs, domains, orthology or the role of key residues, the function of an unknown protein may be unravelled (Zarembinski *et al.*, 1998; Shin *et al.*, 2002; Pal & Eisenberg, 2005). A structural genomics program for a local plant pathogen *Xanthomonas campestris* pv. *campestris* strain 17 (Xcc) has been initiated recently in order to study the structures and functions of unknown genes in the Xcc genome. In the past, genes of unknown function have chiefly been annotated by searching protein or DNA databases for sequence similarities using popular programs such as *BLAST* and *PSI-BLAST* (Altschul *et al.*, 1997). When sequence similarity above 30% is detected, it is likely that the unknown protein will exhibit a similar function. Unfortunately, not every function of unknown genes can be predicted in this way. Many ORFs (open reading frames) exhibit no sequence similarity above 30% and hence their role and function remain unassigned. For example, 1474 of the 4182 annotated ORFs in the published *X. campestris* pv. *campestris* strain ATCC 33913 genome have no assigned function, including 1276 so-called conserved hypothetical proteins (also found in other bacteria) and 198 so-called hypothetical proteins that are detected exclusively in the Xcc genome (da Silva *et al.*, 2002). The functions of these proteins therefore need to be determined by a different approach.

XC1692 has been classified as a 141-amino-acid conserved hypothetical protein using a bioinformatics approach (http://xcc.life.nthu.edu.tw/). It shares 100% identity with a similar protein (gi|21113420) from another strain of the *Xanthomonas* genus (*X. campestris* pv. *campestris* strain ATCC 33913; da Silva *et al.*, 2002), 50% identity with a protein from *Xylella fastidiosa* (Simpson *et al.*, 2000) and 32% identity with a protein from *Ralstonia solanacearum* (Salanoubat *et al.*, 2002) species. To date, no tertiary structure has been reported in the PDB for any protein similar to XC1692 and it has not been classified in any superfamily in the COG (Tatusov *et al.*, 2001) or Pfam databases (Bateman *et al.*, 2000). In this report, we

describe the cloning, purification, crystallization and initial X-ray analyses of XC1692.

2. Materials and methods

2.1. Cloning, expression and purification

The XC1692 gene fragment was PCR-amplified directly from a local Xcc genome. It was cut with the *Ssp*I restriction enzyme and cloned into a ligation-independent vector pTBSG1 (F. P. Gao, unpublished results). The final construct codes for an N-terminal His_6 tag, a 17-amino-acid linker and an XC1692 protein (141 amino acids) under the control of a T7 promoter. Transformed *Escherichia coli* BL21 (DE3) host cells were grown in LB medium at 310 K until an OD_{600} of 0.9 was attained. Overexpression of the His_6-tagged protein was induced by the addition of 0.5 m*M* IPTG at 310 K for 4 h. The cells were harvested, resuspended in equilibration buffer (20 m*M* Tris, 70 m*M* NaCl pH 8.0) and lysed using a microfluidizer (Microfluidics). Most tagged target proteins were in the soluble fraction (Fig. 1). After centrifugation, the crude extract was applied onto a nickel column (Sigma), which was washed with 100 ml 9 m*M* imidazole to remove non-specific protein adsorption. On-column cleavage of the linker-fused target protein was then performed by incubating the nickel column with 20 ml tobacco etch virus (TEV) protease buffer (10% glycerol, 50 m*M* sodium phosphate pH 8.0, 0.5 *M* NaCl). Subsequently, approximately 100 µl 5 mg ml^{-1} TEV protease was added to the column to execute the linker-cleavage reaction. The column was shaken overnight and the free target protein was collected in the flowthrough fractions, with the linker residues and TEV protease retained on the column owing to their His_6 tag. The purified protein was then dialyzed again against 20 m*M* Tris pH 8.0 and 70 m*M* NaCl. For crystallization, XC1692 was further purified on an anion-exchange column (AKTA, Pharmacia Inc.). The fractions eluted with 20 m*M* Tris pH 8.0, 650 m*M* NaCl were combined and dialyzed against 20 m*M* Tris pH 8.0 and 70 m*M* NaCl. The final target protein (141 amino acids) has greater than 99% purity (Fig. 1) and only contains an extra tripeptide (SNA) at the N-terminal end. The

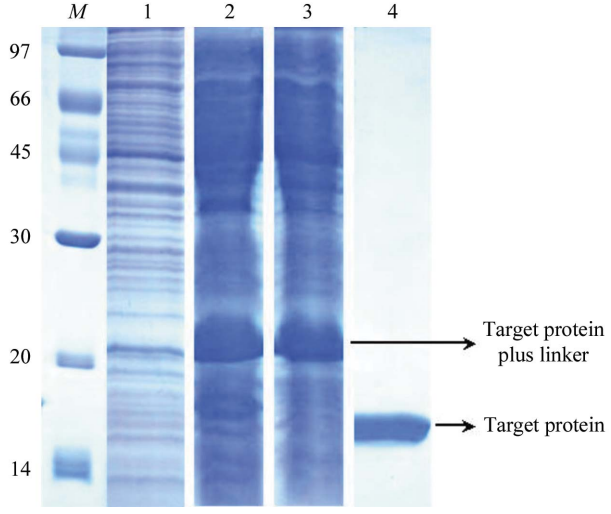


Figure 1
SDS–PAGE monitoring of the overexpression and purification of XC1692 (sample reduced using 100 m*M* dithiothreitol). Lane *M*, molecular-weight markers in kDa; lane 1, soluble fraction before IPTG induction; lane 2, total fraction after IPTG induction; lane 3, soluble fraction after IPTG induction; lane 4, purified XC1692 after TEV protease cleavage. The positions of linker-fused and free target proteins are marked by arrows.

Table 1
Data-collection statistics for XC1692.

Values in parentheses are for the highest resolution shell.

Space group	$P6_3$
Unit-cell parameters (Å)	$a = b = 56.9$, $c = 71.0$
Temperature (K)	100
Wavelength (Å)	1.5418
Resolution range (Å)	49.27–1.45 (1.50–1.45)
Mosaicity (°)	0.6
Unique reflections	22893 (2086)
Redundancy	4.8 (2.3)
Completeness (%)	98.5 (90.8)
R_{merge} (%)	5.1 (43.6)
Mean $I/\sigma(I)$	24.7 (2.5)
Solvent content (%)	43.93

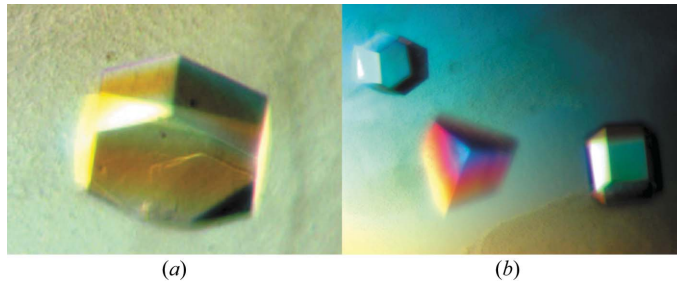
overexpression and purification of XC1692 was monitored by SDS–PAGE as shown in Fig. 1.

2.2. Crystallization

For crystallization purposes, the protein was concentrated to 47 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 condition 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; drops were equilibrated against 50 µl mother liquor. 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. Hexagonal-shaped or prism-shaped crystals appeared in 1 d from a reservoir solution comprising 0.1 *M* HEPES buffer pH 7.5, 0.2 *M* calcium acetate and 15%(*w*/*v*) PEG 4K. Crystallization conditions were further fine-tuned by adjusting buffer types and PEG concentrations. 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 0.4 × 0.4 × 0.3 mm after 3 d (Fig. 2).

2.3. Data collection

Crystals were soaked in a cryoprotectant solution consisting of reservoir solution containing 20%(*v*/*v*) glycerol and were then flash-cooled at 100 K in a stream of cold nitrogen. X-ray diffraction data were collected using Cu *Kα* radiation from a Rigaku RU-300 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 1.45 Å. The data were indexed and integrated



(*a*) (*b*)

Figure 2
Crystals of XC1692 grown by the hanging-drop vapour-diffusion method under different conditions. Those grown in 0.1 *M* Tris buffer pH 8.5, 0.2 *M* calcium acetate, 6%(*w*/*v*) PEG 20K and 8%(*w*/*v*) PEG 550 MME (monomethyl ether) are shown in (*a*), while those grown in 0.1 *M* HEPES buffer pH 7.5, 0.2 *M* calcium acetate and 15%(*w*/*v*) PEG 4K are shown in (*b*). Crystal polymorphism (hexagonal-shaped and prism-shaped) was observed under these conditions. The approximate dimensions of these crystals reached 0.4 × 0.4 × 0.3 mm in 3 d.

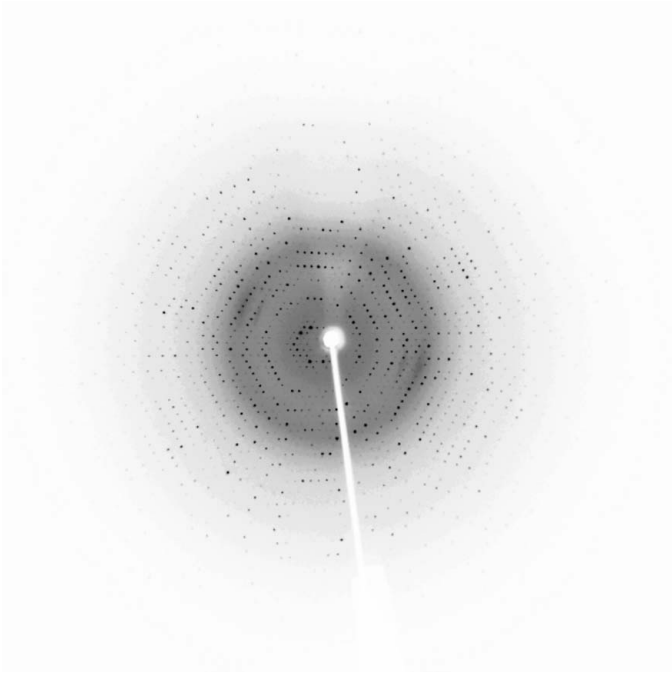


Figure 3
The diffraction pattern obtained in-house from an XC1692 crystal cryoprotected with 20% glycerol and flash-cooled. The exposure time was 10 min, with an oscillation range of 1.0° and a crystal-to-detector distance of 80 mm.

using the *HKL* software suite (Otwinowski & Minor, 1997), giving a 98.5% complete data set with an overall $R_{\rm merge}$ of 5.1% on intensities. The crystals belong to the hexagonal space group $P6_3$, with one molecule in the asymmetric unit, 43.93% solvent content and a $V_{\rm M}$ value of 2.21 Å^3 Da^{-1}. The data-collection statistics are summarized in Table 1. An X-ray diffraction image of an XC1692 protein crystal collected in-house is shown in Fig. 3.

3. Results and discussion

The gene sequence of XC1692 was confirmed after cloning and consists of 423 bp, coding for a protein of 141 amino-acid residues with a calculated isoelectric point of 4.49. The purified XC1692 showed a single band of approximately 15 kDa on SDS–PAGE, with a purity greater than 99% (Fig. 1), which possibly accounts for its straightforward crystallization. Good-quality crystals of dimensions 0.4 × 0.4 × 0.3 mm diffracting to at least 1.45 Å were readily obtained in 3 d (Fig. 2).

We have chosen proteins with unknown structure and/or unknown functions as our targets in order to increase the possibility of discovering novel protein folds. High-resolution preliminary X-ray diffraction data have been obtained for native XC1692 crystals (Fig. 3), indicating their suitability for further detailed X-ray structural analysis. As XC1692 contains two cysteines and three methionines, we plan to solve the structure using either the multiple isomorphous replacement (MIR) method (Ke, 1997) or the multi-wavelength anomalous diffraction (MAD) method using selenomethionine-substituted protein (Hendrickson & Ogata, 1997).

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 ofthe Academia Sinica, Taiwan for assistance in preliminary X-ray analysis.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. L. (2000). *Nucleic Acids Res.* **28**, 263–266.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
Ke, H. (1997). *Methods Enzymol.* **276**, 448–461.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pal, D. & Eisenberg, D. (2005). *Structure*, **13**, 121–130.
Salanoubat, M. *et al.* (2002). *Nature (London)*, **415**, 497–502.
Shin, D. H., Yokota, H., Kim, R. & Kim, S.-H. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 7980–7985.
Silva, A. C. R. da *et al.* (2002). *Nature (London)*, **417**, 459–463.
Simpson, A. J. *et al.* (2000). *Nature (London)*, **406**, 151–156.
Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D. & Koonin, E. V. (2001). *Nucleic Acids Res.* **29**, 22–28.
Zarembinski, T. I., Hung, L.-W., Mueller-Dieckmann, H.-J., Kim, K.-K., Yokota, H., Kim, R. & Kim, S.-H. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 15189–15193.