

**Ko-Hsin Chin,<sup>a</sup> Wei-Tien Kuo,<sup>a</sup>  
 Chia-Cheng Chou,<sup>b,c</sup> Hui-Lin  
 Shr,<sup>b,c</sup> Ping-Chiang Lyu,<sup>d</sup>  
 Andrew H.-J. Wang<sup>b,c</sup> and  
 Shan-Ho Chou<sup>a\*</sup>**

<sup>a</sup>Institute of Biochemistry, National Chung-Hsing University, Taichung 40227, Taiwan,

<sup>b</sup>Institute of Biological Chemistry, Academia Sinica, Nankang, Taipei, Taiwan, <sup>c</sup>Core Facility for Protein Crystallography, Academia Sinica, Nankang, Taipei, Taiwan, and <sup>d</sup>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

## Cloning, purification, crystallization and preliminary X-ray analysis of XC229, a conserved hypothetical protein from *Xanthomonas campestris*

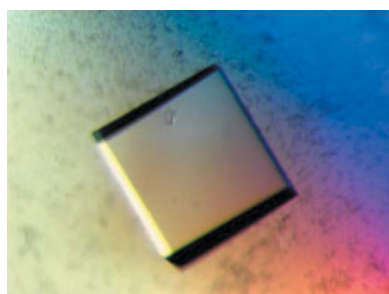
*Xanthomonas campestris* pv. *campestris* is a Gram-negative yellow-pigmented pathogenic bacterium that causes black rot, one of the major worldwide diseases of cruciferous crops. Its genome contains approximately 4500 genes, roughly one third of which have no known structure and/or function. However, some of these unknown genes are highly conserved among several different bacterial genuses. XC229 is one such protein containing 134 amino acids. It was overexpressed in *Escherichia coli*, purified and crystallized using the hanging-drop vapour-diffusion method. The crystal diffracted to a resolution of at least 1.80 Å. It is cubic and belongs to space group  $I2_x3$ , with unit-cell parameters  $a = b = c = 106.8$  Å. It contains one or two molecules per asymmetric unit.

### 1. Introduction

A structural genomics program for a local plant pathogen *Xanthomonas campestris* pv. *campestris* strain 17 (Xcc) has recently been initiated 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 identity above 30%. Hence, many of them remain unassigned. For example, of the 4182 annotated ORFs in the published *X. campestris* pv. *campestris* strain ATCC 33913 genome, 1474 have no assigned function, including 1276 so-called conserved hypothetical proteins that are also found in other bacteria and 198 so-called hypothetical proteins that are only detected in the Xcc genome (da Silva *et al.*, 2002). The functions of such proteins therefore need to be elucidated by a different approach.

One of the major goals of a structural genomics program is to produce as many three-dimensional protein structures as possible in order to provide a better understanding of the protein sequence–structure–function relationship (Zarembinski *et al.*, 1998; Shin *et al.*, 2002; Pal & Eisenberg, 2005). From structural information on fold, motifs, domains, orthology or functionally significant residues, the functions of an unknown protein may be unravelled.

XC229 (gi|21112185) has been classified as a conserved hypothetical protein from sequence comparison (<http://xcc.life.nthu.edu.tw/>). It contains 134 amino acids and shares 40% identity with a protein from *Pseudomonas aeruginosa* (gi|15597997; Stover *et al.*, 2000), 29% identity with a protein from *Ralstonia solanacearum* (gi|17427742; Salanoubat *et al.*, 2002) and 100% identity with a protein from *Xanthomonas campestris* pv. *campestris* strain ATCC33913 (gi|21112185; da Silva *et al.*, 2002), respectively. To date, no tertiary structure has been reported in the PDB for any protein similar to XC229, although it is classified in the putative thioesterase superfamily in the Pfam database (Bateman *et al.*, 2000), which contains a wide variety of enzymes. This family includes various cytosolic long-chain acyl-CoA thioester hydrolases that catalyse the hydrolysis of several long-chain fatty acyl-CoA thioesters. In this report, we describe the cloning, purification, crystallization and initial X-ray analyses of XC229.

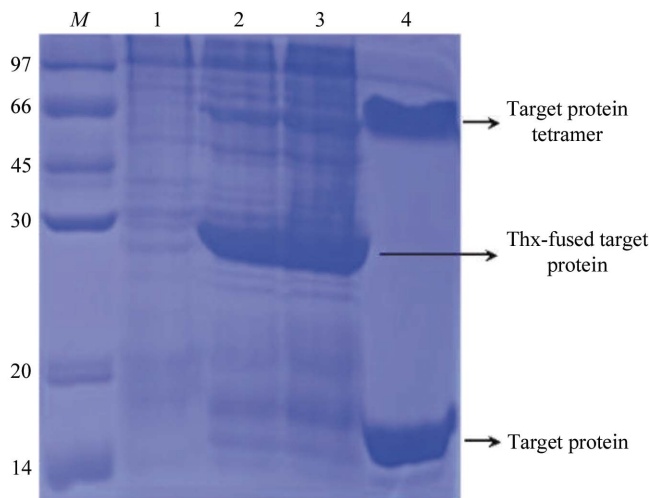


© 2005 International Union of Crystallography  
 All rights reserved

## 2. Materials and methods

### 2.1. Cloning, expression and purification

The XC229 gene fragment was PCR-amplified directly from a local *Xcc* genome (*X. campestris* pv. *campestris* strain 17), which shares greater than 99.5% identity with the *Xcc* strain ATCC33913 published previously (da Silva *et al.*, 2002). It was cut with *EcoRI* and *XhoI* restriction enzymes and cloned into a modified pET-32a(+) vector (Shih *et al.*, 2002). The final construct codes for a thioredoxin tag protein (109 amino acids), XC229 protein (134 amino acids) and a C-terminal His<sub>6</sub> tag under the control of a T7 promoter. The transformed *Escherichia 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 mM IPTG at 293 K for 20 h. The cells were harvested, resuspended in equilibration buffer (20 mM Na<sub>2</sub>HPO<sub>4</sub>, 70 mM NaCl pH 8.5) and lysed using a microfluidizer (Microfluidics). After centrifugation, the tagged protein was purified by immobilized metal-affinity chromatography (IMAC) on a cobalt column (BD Biosciences). The fusion protein was then eluted with 20 mM Tris pH 8.0, 70 mM NaCl and a gradient of 100–300 mM imidazole. The fractions containing XC229 were dialyzed repeatedly with 140 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub> and 1.8 mM KH<sub>2</sub>PO<sub>4</sub>. After concentration, the mixture was loaded onto a cobalt column and the thioredoxin tag removed from the bound fusion protein by thrombin cleavage at 295 K for 16 h. The XC229 target protein was then eluted with 20 mM Tris pH 8.0, 100–300 mM imidazole and 70 mM NaCl. The desired fractions were dialyzed repeatedly in 20 mM Tris pH 8.0 and 70 mM NaCl. For crystallization, the XC229 protein was further purified on an anion-exchange column (Pharmacia Inc.). The fractions eluted with 20 mM Tris pH 8.0, 500 mM NaCl were combined and dialyzed against 20 mM Tris pH 8.0 and 70 mM NaCl. The final construct (150 amino acids) contains the target protein (134 amino acids), an extra octapeptide (GSGGGGEF) at the N-terminal end and an extra octapeptide LEH<sub>6</sub> at the C-terminal end, with an MW of 16621 Da consistent with mass-spectrometric data. The overexpression and purification of XC229 was monitored on SDS-PAGE as shown in Fig. 1.



**Figure 1**

SDS-PAGE monitoring of the overexpression and purification of XC229. Lane M, molecular-weight markers 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, purified XC229 after thrombin cleavage. The positions of target protein tetramer, Thx-fused target protein and free target protein are also marked.

**Table 1**

Data-collection statistics for XC229.

Values in parentheses are for the highest resolution shell.

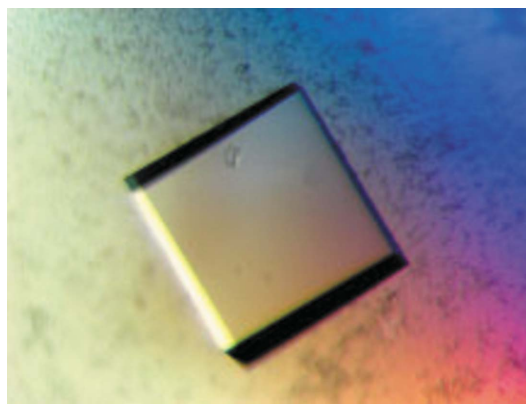
Space group	<i>I</i> 2 <sub>1</sub> 3
Unit-cell parameters (Å)	<i>a</i> = <i>b</i> = <i>c</i> = 106.8
Temperature (K)	100
Wavelength (Å)	1.5418
Resolution range (Å)	39.3–1.80 (1.86–1.80)
Unique reflections	18909 (1890)
Redundancy	6.4 (5.9)
Mosaicity (°)	0.2
Completeness (%)	99.9 (100.0)
<i>R</i> <sub>merge</sub> (%)	3.7 (35.4)
Mean <i>I</i> σ( <i>I</i> )	28.1 (4.6)
Solvent content (%)	63.59 or 27.18
Cryoprotectant	Mother liquor

### 2.2. Crystallization

For crystallization, the protein was concentrated to 55 mg ml<sup>-1</sup> in 20 mM Tris pH 8.0 and 70 mM 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 290 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. Cubic-shaped crystals appeared within 1 d from a reservoir solution comprising 37% MPD (2-methyl-2,4-pentanediol). This initial condition was then optimized by varying the concentration of MPD. Crystals suitable for diffraction experiments were grown by mixing 1.5 μl protein solution with 1.5 μl reagent solution containing 32% MPD using the hanging-drop vapour-diffusion method and reached maximum dimensions of 1.0 × 1.0 × 0.3 mm after one week (Fig. 2).

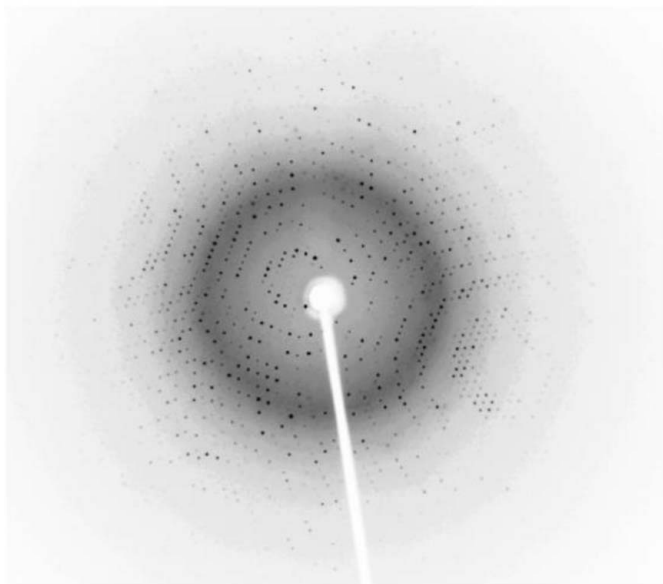
### 2.3. Data collection

Crystals were soaked in a cryoprotectant solution containing 25% glycerol in addition to the components of the reservoir solution and were flash-cooled at 100 K in a stream of cold nitrogen. X-ray diffraction data were collected using Cu *K*α radiation from a Rigaku MicroMax007 rotating-anode generator equipped with Osmic mirror optics and an R-Axis IV<sup>++</sup> image plate. A native data set was collected to a maximum resolution of 1.8 Å. The data were indexed and integrated using the *HKL* software suite (Otwinowski & Minor,



**Figure 2**

Crystallization of XC229 from *X. campestris*. Crystals of XC229 were grown by the hanging-drop vapour-diffusion method under the final optimized crystallization conditions of 32% MPD. The approximate dimensions of these crystals were 1.0 × 1.0 × 0.3 mm after one week.



**Figure 3**

Picture of the diffraction pattern of XC229 collected in-house from a flash-frozen crystal in mother-liquor cryoprotectant. The exposure time was 15 min, with an oscillation range of  $1.0^\circ$  and a crystal-to-detector distance of 110 mm.

1997), giving a data set that was 99.9% complete with an overall  $R_{\text{merge}}$  of 4.1% on intensities. The crystals belong to the cubic space group  $I2_13$ , with one or two molecules in the asymmetric unit and 63.6 or 27.2% 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 XC229 was confirmed after cloning and consists of 402 bp coding for 134 amino-acid residues. The purified XC229 contains an extra octapeptide (GSGGGGEF) at the N-terminal end and an extra octapeptide LEH<sub>6</sub> at the C-terminal end and is greater than 97% pure, with a single band of approximately 16.6 kDa on SDS-PAGE (Fig. 1). Interestingly, an extra band of approximately 66 kDa was also observed (Fig. 1), indicating that a stable tetramer is in equilibrium with the monomer even under the denatured conditions used in the PAGE analysis. This conclusion is consistent with the fact that when the PAGE experiment was carried out with sample first heated at 363 K for 5 min before loading, only one band, corresponding to the monomer, was observed (data not

shown). The possibility that XC229 forms a tetramer owing to disulfide-bond formation between monomers can be eliminated since it contains no cysteines. The extra 16 amino acids at the ends and the heterogeneous mixture of monomer and tetramer do not seem to affect the crystallization process, as crystals readily grew to dimensions of  $1.0 \times 1.0 \times 0.3$  mm (Fig. 2) overnight and diffracted to a good resolution of at least  $1.80 \text{ \AA}$  (Fig. 3).

We have chosen proteins with unknown structure and/or unknown functions as our targets for this structural genomics project in order to increase the possibility of discovering novel protein folds and have so far obtained many good crystals using these targets. XC229 was found to give good diffraction data suitable for further detailed X-ray structural analysis (Fig. 3). We now plan to solve the structure of XC229 using either the multiple isomorphous replacement (MIR) method by preparing platinum or gold heavy-atom derivatives (Ke, 1997) or by the multiwavelength anomalous diffraction (MAD) method using selenomethionine-substituted protein (Hendrickson & Ogata, 1997), since a single XC229 contains five methionines. Heavy-atom positions and phases will be determined using automated Patterson analysis as described by Terwilliger & Berendzen (1999).

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

- 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.
- 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.
- 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.
- Stover, C. K. *et al.* (2000). *Nature (London)*, **406**, 959–964.
- Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* **D55**, 849–861.
- 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.