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

# Preparation, crystallization and preliminary X-ray analysis of XC2382, an ApaG protein of unknown structure from *Xanthomonas campestris*

**Ko-Hsin Chin,[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* is the causative agent of black rot, one of the major worldwide diseases of cruciferous crops. Its genome encodes approximately 4500 proteins, roughly one third of which have unknown function. XC2382 is one such protein, with a MW of 14.2 kDa. Based on a bioinformatics study, it was annotated as an ApaG gene product that serves multiple functions. The ApaG protein has been overexpressed in *Escherichia coli*, purified and crystallized using the hanging-drop vapour-diffusion method. The crystals diffracted to a resolution of at least 2.30 Å. They are tetragonal and belong to space group $P4_{1/3}$, with unit-cell parameters $a = b = 57.6$, $c = 122.9$ Å. There are two, three or four molecules in the asymmetric unit.

## 1. Introduction

Structural genomics aims to determine the three-dimensional structures of all proteins on a genomic scale (Edwards *et al.*, 2004; Zhang & Kim, 2004). We have initiated a structural genomics program focused on *Xanthomonas campestris* pv. *campestris* strain 17 (Xcc), a phytopathogen that is widespread and causes diseases of cruciferous crops (da Silva *et al.*, 2002). We are trying to understand in structural terms its pathogenicity toward its host (Dow *et al.*, 2000; da Silva *et al.*, 2002), its capability to produce xanthan gum (Tseng *et al.*, 1999), its protein-secretion pathway (Lee *et al.*, 2001), its multiple gene-regulation pathways (Ebright, 1993) and its particular response under stress (Santos *et al.*, 1999).

The Xcc genome encodes approximately 4500 proteins, many of which have unknown function. One way to explore the function of an unknown gene is to determine how it is connected to other genes that exhibit known functions. Hence, when an unknown gene belongs to the same operon as another known gene, the function of the unknown gene can be partially deduced, especially when its three-dimensional structural information is available. The gene encoding XC2382, a protein with a molecular weight of 14.2 kDa, has been named *apa*G by a bioinformatics study and is located in a multifunctional *ksg*A-*apa*G-*apa*H operon (Blachin-Roland *et al.*, 1986). The expression of *apa*G is found to be tightly linked to *apa*H, a gene whose expression product is involved in the metabolism of $Ap_4N$ (Leveque *et al.*, 1990). $Ap_4Ns$ are unusual tetraphosphate nucleotides found to be responsible for the priming of DNA replication or to serve as intermediates in the synthesis of alarmones to signal the outset of cellular stress. ApaG protein may thus serve a function in the homeostasis of these interesting metabolites. However, as ApaG has also been suggested to possess other functions (Rossmann *et al.*, 1974; Wierenga *et al.*, 1986; Llyin *et al.*, 2000; Tatusov *et al.*, 2001; Liu *et al.*, 2003), its precise role remains a mystery. To further investigate the function of XC2382, we have engaged in the solution of its structure. We have cloned this gene into the *Escherichia coli* host and expressed, purified and crystallized its gene product. Initial X-ray analysis of XC2382 was also performed and is described in this report.

## 2. Materials and methods

### 2.1. Cloning, expression and purification

The XC2382 gene was amplified using the sticky PCR method (Zeng, 1998) from a local Xcc strain. Two PCR reactions, one initi-

ated with the forward primer P1-**AATTC**GGAATGCAAGATGATCCGCGCTACC and the backward primer P2-**G**GTGCAGCGTCCTCGGCACGCTC, and the other with the forward primer P3-**C**GGAATGCAAGATGATCCGCGCTACC and the backward primer P4-**TCGAG**GTGCAGCGTCCTCGGCAC, 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 maltose-binding protein tag (396 amino acids), an XC2382 protein (127 amino acids) and a C-terminal $His_6$ tag under the control of a T7 promoter. Transformed *E. coli* BL21 (DE3) host cells were 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 310 K for 4 h. The cells were harvested, resuspended in equilibration buffer (20 m*M* $Na_2HPO_4$, 70 m*M* NaCl pH 8.0) and lysed using a microfluidizer (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 XC2382 were dialyzed repeatedly with 50 m*M* Tris pH 8.0, 100 m*M* NaCl and 5 m*M* CaCl. After concentration, the mixture was loaded onto a cobalt column and the maltose-binding protein (MBP) tag was removed from the bound fusion protein by factor Xa cleavage at 295 K for 16 h. The XC2382 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 against 20 m*M* Tris pH 8.0 and 70 m*M* NaCl. For crystallization, XC2382 was further purified on an anion-exchange column (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 (127 amino acids) contains an extra octapeptide (GSGGGGEF) at the N-terminal end and an extra octapeptide $LEH_6$ at the C-terminal end, with a molecular weight of 15.8 kDa, which was confirmed by mass-spectrometric analysis. The overexpression and purification of XC2382 was monitored on SDS–PAGE as shown in Fig. 1.

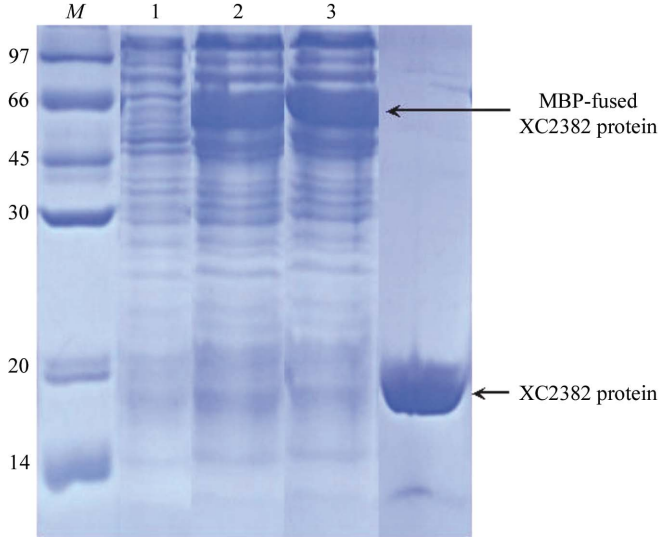


**Figure 1**
SDS–PAGE monitoring of the overexpression and purification of XC2382. Lane *M*, molecular-weight markers in kDa; lane 1, whole cell lysate before IPTG induction; lane 2, whole cell lysate after IPTG induction; lane 3, soluble fraction after IPTG induction; lane 4, purified XC2382 after factor Xa cleavage. The positions of MBP-fused XC2382 (59.4 kDa) and XC2382 with linker residues (15.8 kDa) are marked by arrows.

**Table 1**
Data-collection statistics for XC2382.

Values in parentheses are for the highest resolution shell.

| | |
|---|---|
| Space group | $P4_{1/3}$ |
| Unit-cell parameters (Å) | $a = b = 57.6$, $c = 122.9$ |
| Temperature (K) | 100 |
| Wavelength (Å) | 1.5418 |
| Resolution range (Å) | 40.70–2.30 (2.38–2.30) |
| Mosaicity (°) | 0.78 |
| No. of unique reflections | 17755 (1764) |
| Redundancy | 3.3 (3.2) |
| Completeness (%) | 99.8 (99.6) |
| $R_{merge}$ (%) | 5.7 (41.1) |
| Mean $I/\sigma(I)$ | 16.4 (3.3) |
| Solvent content (%) | 65.5, 48.2 or 30.9 |

### 2.2. Crystallization

For crystallization, the protein was concentrated to 32 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. Pillar-shaped crystals appeared in 1 d from a reservoir solution comprising 0.8 *M* $NaH_2PO_4$ pH 7.0 and 0.8 *M* $K_2HPO_4$ pH 7.0. This initial conditions were then optimized by varying the concentrations of sodium phosphate and potassium phosphate. 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 1.2 × 0.2 × 0.1 mm after 3 d (Fig. 2).

### 2.3. Data collection

Crystals were first frozen in liquid nitrogen and then gradually soaked in a cryoprotectant solution containing 15, 20 and 22%(*v*/*v*) glycerol in addition to the components of the reservoir. They 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

**Figure 2**
Crystallization of ApaG from *X. campestris* obtained by the sitting-drop vapour-diffusion method. The final applied crystallization condition was 0.6 *M* sodium phosphate/potassium phosphate pH 7.0. The approximate dimensions of these crystals reached 1.2 × 0.2 × 0.1 mm after 3 d.

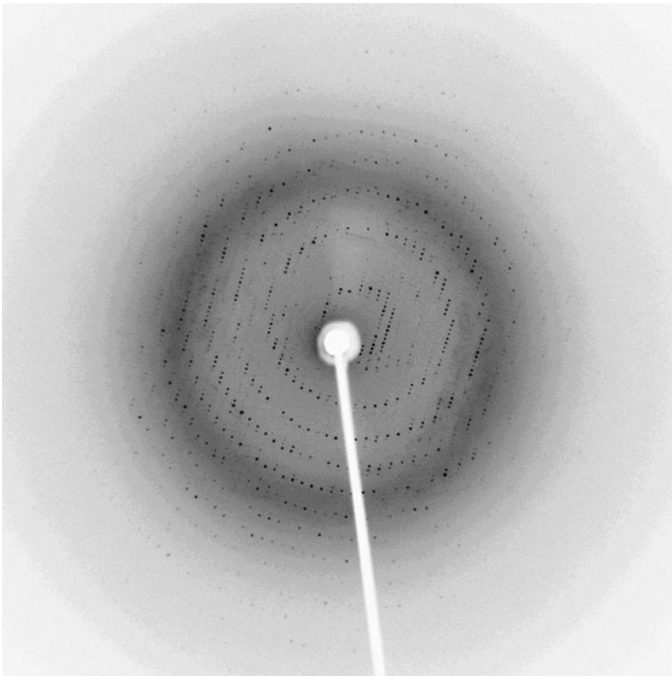

**Figure 3**
The diffraction pattern collected in-house from a flash-frozen crystal with 22% glycerol cryoprotectant. The exposure time was 10 min, with an oscillation range of 1.0° and a crystal-to-detector distance of 150 mm.

RU-300 rotating-anode generator equipped with Osmic mirror optics and an R-AXIS IV$^{++}$ image plate. A resolution of only 2.3 Å could be obtained for the native data set because of the limitation of a long unit-cell parameter. The data were indexed and integrated using *HKL* processing software (Otwinowski & Minor, 1997), giving a data set that was 99.8% complete with an overall $R_{\rm merge}$ of 5.7% on intensities. The crystals belong to the tetragonal space group $P4_{1/3}$, with two, three or four molecules in the asymmetric unit, corresponding to 65.5, 48.2, or 30.9% 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

XC 2382 was predicted to be an ApaG protein (gi|21111809) based on a bioinformatics study (http://xcc.life.nthu.edu.tw/). It bears 74% sequence identity to the ApaG protein (gi|15838740) from *Xylella fastidiosa* (Simpson *et al.*, 2000), 100% identity to the ApaG protein (gi|21111809) from Xcc ATCC 33913 (da Silva *et al.*, 2002) and 51% identity to a protein of unknown function (gi|17429901) from *Ralstonia solanacearum* (Salanoubat *et al.*, 2002). It contains a highly conserved G*X*G*XX*G context sequence, a pyrophosphate-binding motif that has been found in the NAD- and FAD-binding proteins (Rossmann *et al.*, 1974; Wierenga *et al.*, 1986), suggesting that it may bind pyrophosphate or nucleotide phosphates. ApaG also shares sequence homology with CorD, a *Salmonella typhimurium* protein associated with $Co^{2+}$ sensitivity and $Mg^{2+}$ homeostasis, and is therefore classified as an uncharacterized protein affecting $Mg^{2+}/Co^{2+}$ transport in the COG database (Tatusov *et al.*, 2001). Furthermore, ApaG-homologue domains have also been found in eukaryotic F-box proteins involved in the recognition of substrate destined for ubiquitination (Llyin *et al.*, 2000; Liu *et al.*, 2003). The high degree of sequence conservation among ApaG homologues in bacteria indicated that it performs some important biological function.

The gene of XC2382 consists of 381 bp coding for 127 amino-acid residues (da Silva *et al.*, 2002). The isoelectric point was calculated to be 5.12. The purified XC2382 contains an extra octapeptide GSGGGGEF at the N-terminal end and an extra octapeptide $LEH_6$ at the C-terminal end and shows a greater than 95% purity, with a single band of approximately 15 kDa on SDS–PAGE (Fig. 1)

The good resolution diffraction obtained from the native crystals establishes their suitability for X-ray structural analysis (Fig. 3). We now plan to solve the structure of XC2382 by the multiwavelength anomalous diffraction (MAD) method using selenomethionine-substituted protein (Hendrickson & Ogata, 1997; Terwilliger & Berendzen, 1999), given that a single XC2382 contains three methionines. However, it may also be possible to solve the XC2382 structure directly by the molecular-replacement method, as an ApaG crystal structure from *Vibrio cholerae* sharing 42% sequence identity and 64% similarity with Xcc ApaG has recently been solved (PDB code 1xvs). Interestingly, an ApaG from *X. axonopodis* pv. *citri* sharing 94% identity with the Xcc ApaG has had its backbone assigned by the NMR method (BMRB5998; Katsuyama *et al.*, 2004). These studies will allow a more thorough understanding of this remarkable protein of diverse functions.

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

Blachin-Roland, S. S., Blanquet, S., Schmitter, J.-M. & Fayat, G. (1986). *Mol. Gen. Genet.* **205**, 515–522.
Dow, J. M., Feng, J.-X., Barber, C. E., Tang, J.-L. & Daniels, M. J. (2000). *Microbiology*, **146**, 885–891.
Ebright, R. H. (1993). *Mol. Microbiol.* **8**, 797–802.
Edwards, A. F., Yee, A., Savchenko, A., Edwards, A. M. & Arrowsmith, C. H. (2004). *Curr. Opin. Chem. Biol.* **8**, 42–48.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
Katsuyama, A. M., Cicero, D. O., Spisni, A., Paci, M., Farah, C. S. & Pertinhez, T. A. (2004). *J. Biomol. NMR*, **29**, 423–424.
Lee, H.-M., Tyan, S.-W., Leu, W.-M., Chen, L.-Y., Chen, D. C. & Hu, N.-T. (2001). *J. Bacteriol.* **183**, 528–535.
Leveque, F., Blanchin-Roland, S., Fayat, G., Plateau, P. & Blanquet, S. (1990). *J. Mol. Biol.* **20**, 319–329.
Liu, L., Rodriguez-Belmonte, M., Mazolum, N., Xie, B. & Lee, M. Y. W. T. (2003). *J. Biol. Chem.* **278**, 10041–10047.
Llyin, G. P., Raialland, M., Pigeon, C. & Guguen-Guillouzo, C. (2000). *Genomics*, **67**, 40–47.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rossmann, M. G., Moras, D. & Olsen, K. W. (1974). *Nature (London)*, **250**, 194–199.
Salanoubat, M. *et al.* (2002). *Nature (London)*, **415**, 497–502.
Santos, J. M., Freire, P., Vicente, M. & Arraiano, C. M. (1999). *Mol. Microbiol.* **32**, 789–798.
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.
Simpson, A. J. G. *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.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Tseng, Y.-H., Choy, K.-T., Hung, C.-H., Lin, N.-T., Liu, J.-Y., Lou, C.-H., Yang, B.-Y., Wen, F.-S., Weng, S.-F. & Wu, J.-R. (1999). *J. Bacteriol.* **181**, 117–125.
Wierenga, R. K., Terpstra, P. & Hol, W. G. (1986). *J. Mol. Biol.* **187**, 101–107.
Zeng, G. (1998). *Biotechniques*, **25**, 206–208.
Zhang, C. & Kim, S.-H. (2004). *Curr. Opin. Chem. Biol.* **7**, 28–32.