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

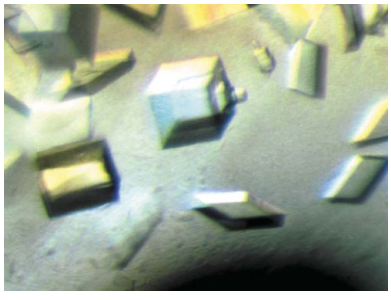
Cloning, crystallization and preliminary X-ray study of XC1258, a CN-hydrolase superfamily protein from *Xanthomonas campestris*

Ying-Der Tsai,[a] Ko-Hsin Chin,[a] 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 10 August 2006
Accepted 2 September 2006



CN-hydrolase superfamily proteins are involved in a wide variety of non-peptide carbon–nitrogen hydrolysis reactions, producing some important natural products such as auxin, biotin, precursors of antibiotics *etc.* These reactions all involve attack on a cyano or carbonyl carbon by a conserved novel catalytic triad Glu-Lys-Cys through a thiol acylenzyme intermediate. However, classification into the CN-hydrolase superfamily based on sequence similarity alone is not straightforward and further structural data are necessary to improve this categorization. Here, the cloning, expression, crystallization and preliminary X-ray analysis of XC1258, a CN-hydrolase superfamily protein from the plant pathogen *Xanthomonas campestris* (Xcc), are reported. The SeMet-substituted XC1258 crystals diffracted to a resolution of 1.73 Å. They are orthorhombic and belong to space group $P2_12_12$, with unit-cell parameters a = 143.8, b = 154.63, c = 51.3 Å, respectively.

1. Introduction

XC1258 (gi|21113352) from the plant pathogen *Xanthomonas campestris* pv. *campestris* strain 17 (Xcc) is classified as belonging to the CN-hydrolase superfamily (previously called the nitrilase superfamily; Pace & Brenner, 2001) in the Pfam database (Bateman *et al.*, 2000). The monomer consists of 266 amino acids and shares sequence identities of 91, 49, 44, 32, 30 and 30% with similar proteins from the *X. campestris* pv. *vesicatoria*, *Pseudomonas putida*, *Erwinia carotovora* subsp. *atroseptica*, *Staphylococcus aureus*, *Bacillus anthracis* and *Lactobacillus salivarius* subsp. *salivarius*, respectively. The CN-hydrolase superfamily is relatively large, containing enzymes such as nitrilases, aliphatic amidases, carbamylase and β-alanine synthase (Pace & Brenner, 2001; Bork & Koonin, 1994). Their reactions usually involve attack on a cyano or carbonyl carbon by a conserved novel catalytic triad Glu-Lys-Cys through a thiol acylenzyme intermediate (Pace & Brenner, 2001).

Currently, only a few tertiary structures are available for proteins belonging to this superfamily, including a NitFhit protein from *Caenorhabditis elegans* (Pace *et al.*, 2000), the *N*-carbamyl-D-amino-acid amidohydrolases (DCases) from *Agrobacterium* sp. strain KNK712 (Nakai *et al.*, 2000) and *A. radiobacter* (Wang *et al.*, 2001) and a putative CN hydrolase from *Sacchromyces cerevisiae* strain S288C (Kumaran *et al.*, 2003), which are either dimeric or tetrameric $\alpha\beta\beta\alpha$ sandwich proteins. Intriguingly, while the structures of *Rhodococcus* nitrilases (nitrile hydratases that convert a nitrile to the corresponding amide; Huang *et al.*, 1997; Nagashima *et al.*, 1998) also comprise $\alpha\beta\beta\alpha$ sandwiches, they are not classified as belonging to the CN-hydrolase superfamily owing to very low structural similarity (Pace & Brenner, 2001). Therefore, more structural studies of proteins from this superfamily are necessary to unambiguously classify this interesting class of proteins. In this manuscript, we describe the cloning, expression, crystallization and preliminary X-ray analysis of XC1258 from *X. campestris*. Good crystals that diffracted to a resolution of 1.73 Å have been obtained. The preliminary structure obtained indicates that the protein adopts a novel quaternary struc-

ture variant of previously determined CN-hydrolase superfamily protein structures.

2. Materials and methods

2.1. Cloning, expression, and purification

The XC1258 gene fragment was PCR amplified directly from a local Xcc genome (*X. campestris* pv. *campestris* strain 17) with forward 5′-TACTTCCAATCCAATGCTATGCACGACCTGCGC-ATTTCCCTC and reverse 5′-TTATCCACTTCCAATGCTACTCG-CCCAGCACAAAGCTGTCCC primers. A ligation-independent cloning (LIC) approach from to a previously published protocol (Wu *et al.*, 2005) was carried out to obtain the desired construct. The final construct codes for an N-terminal His_6 tag, a 17-amino-acid linker and an XC1258 target protein (266 amino acids) under the control of a T7 promoter. Overexpression of the His_6-tagged target protein was induced by the addition of 1 m*M* IPTG at 293 K for 20 h. The target protein was purified by immobilized metal-affinity chromatography (IMAC) on a nickel column (Sigma). The His_6 tag and linker were cleaved from XC1258 by TEV (tobacco etch virus) protease at 288 K for 24 h. For crystallization, XC1258 was further purified on an anion-exchange column (AKTA, Pharmacia Inc.). The final target protein (266 amino acids) has greater than 99% purity (Fig. 1) and contains only an extra tripeptide (SNA) at the N-terminal end. The over-expression and purification of XC1258 was monitored by SDS–PAGE as shown in Fig. 1.

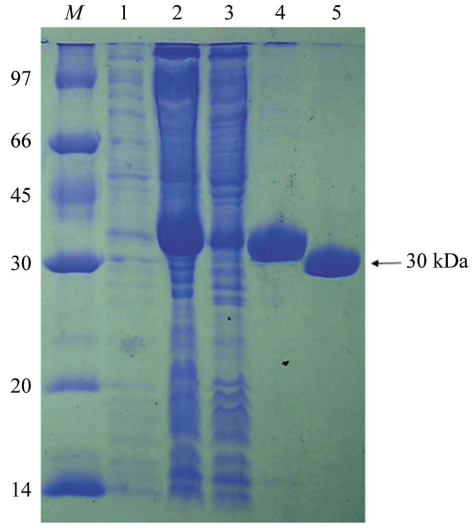


Figure 1
SDS–PAGE monitoring of the overexpression and purification of XC1258. Lane *M*, molecular-weight markers in kDa; lane 1, whole cell lysate before IPTG induction; lane 2, whole cell lysate after IPTG induction; lane 3, soluble fraction after IPTG induction; lane 4, purified XC1258 before TEV cleavage; lane 5, purified XC1258 after TEV cleavage.

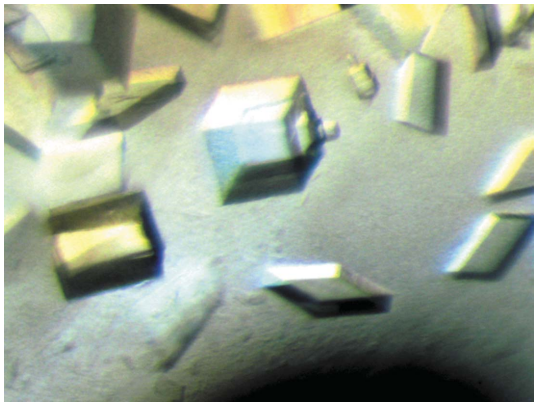


Figure 2
Cube-shaped crystals of XC2981 from *X. campestris* grown by the hanging-drop vapour-diffusion method. The crystallization condition used was 0.1 *M* cacodylate pH 6.5 and 1 *M* sodium citrate. The average dimensions of these crystals were around 0.15 × 0.15 × 0.1 mm.

Table 1
Data-collection statistics for XC1258.

Values in parentheses are for the highest resolution shell.

Space group	$P2_12_12$
Unit-cell parameters (Å)	a = 143.8, b = 154.63, c = 51.3
Temperature (K)	100
Wavelength (Å)	0.96108
Resolution range (Å)	30.0–1.73 (1.79–1.73)
Mosaicity (°)	0.4
Unique reflections	731068 (112277)
Redundancy	6.5 (5.9)
Completeness (%)	98.4 (95.3)
R_{merge} (%)	4.3 (13.5)
Mean $I/\sigma(I)$	33.1 (13.1)
Solvent content (%)	48.5

2.2. Crystallization

For crystallization, the protein was concentrated to 28 mg ml^{-1} in 20 m*M* Tris pH 8.0 and 80 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, including the Hampton Research sparse-matrix Crystal Screens 1 and 2, a systematic PEG–pH screen and a PEG/Ion Screen were performed using a Gilson C240 crystallization workstation. Cube-shaped crystals appeared in 3 d from a reservoir solution comprising 0.1 *M* sodium cacodylate pH 6.5 and

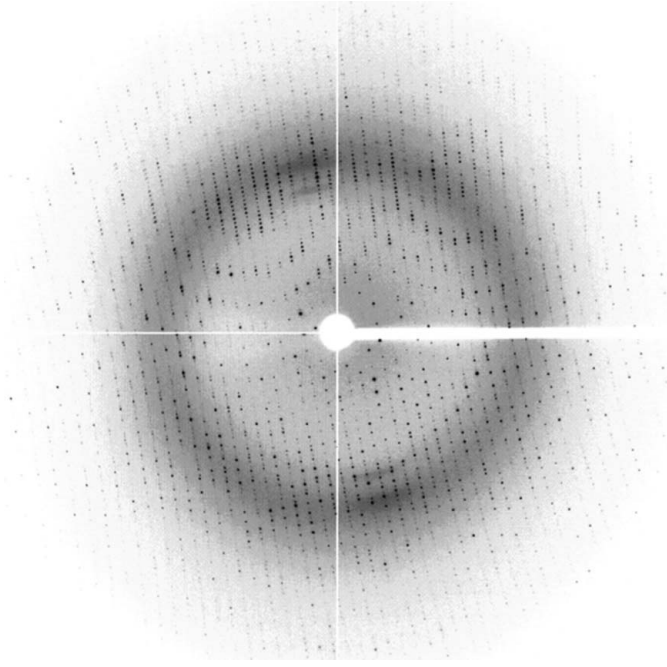


Figure 3
Diffraction pattern of Se-XC1258 collected at NSRRC beamline 13B1 from a crystal flash-frozen in reservoir solution. The exposure time was 3 s, with an oscillation range of 1° and a crystal-to-film distance of 150 mm.

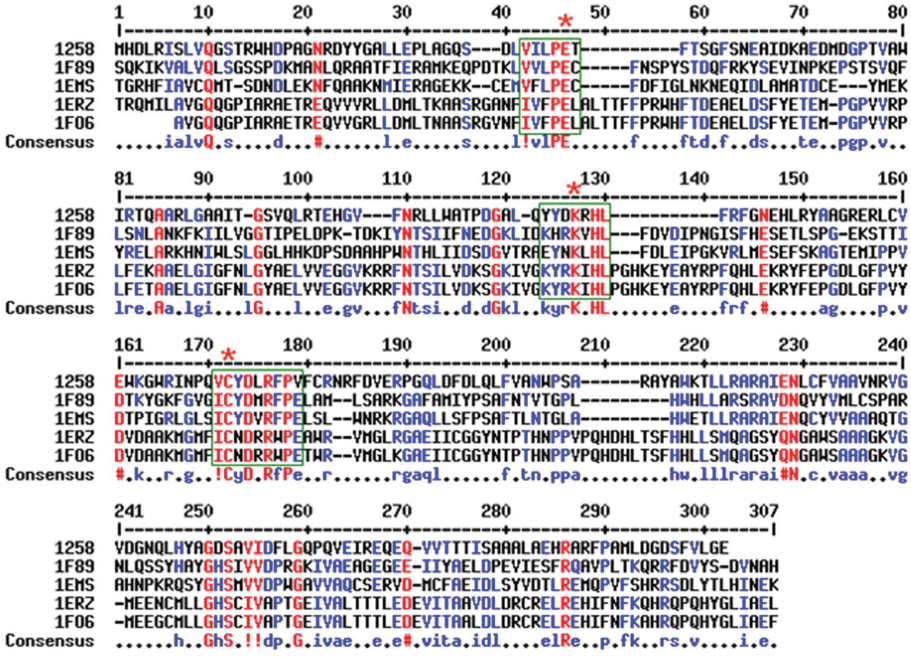


Figure 4
Sequence alignment of XC1258, yeast CN hydrolase (1F89), NitFhit from *C. elegans* (1EMS), DCase from *Agrobacterium* sp. strain KNK712 (1ERZ) and DCase from *A. radiobacter* (1FO6) by *MultiAlign* (Corpet, 1988). Strictly conserved residues are in red and highly conserved residues are in blue. The CN-hydrolase superfamily catalytic triad motifs are boxed in green, with the triad Glu, Lys and Cys residues marked with purple stars.

1 *M* sodium citrate. Crystals suitable for diffraction experiments were grown by mixing 1.5 µl protein solution with 1.5 µl reagent solution at 277 K and reached maximum dimensions of 0.15 × 0.15 × 0.1 mm after one week (Fig. 2).

2.3. Data collection

Crystals were flash-cooled at 100 K without cryoprotectant in a stream of cold nitrogen to prevent crystal cracking. X-ray diffraction data were collected using the National Synchrotron Radiation Research Center (NSRRC) beamline 13B1 in Taiwan. Only a faint ice ring was observed under such circumstances and a data set to 1.73 Å resolution could be obtained for SeMet-substituted XC1258. The data were indexed and integrated using the *HKL*-2000 software (Otwinowski & Minor, 1997), giving a data set that was 98.4% complete with an overall $R_{\rm merge}$ of 4.3%. The crystals belong to the orthorhombic space group $P2_12_12$. The data-collection statistics are summarized in Table 1; an X-ray diffraction image collected at the NSRRC facility is shown in Fig. 3.

3. Results and discussion

The XC1258 gene consists of 801 bp coding for 266 amino-acid residues. The isoelectric point was calculated to be 5.83 using the *Compute pI/MW* tool at the ExPASy website (http://us.expasy.org/tools/pi_tool.html). Purified XC1258 protein shows a single band at the 30 kDa position on SDS–PAGE (Fig. 1), close to its monomer MW.

The high-resolution diffraction data obtained from the SeMet-substituted XC1258 crystals established their suitability for X-ray structural analysis (Fig. 3). A three-wavelength data set has been collected at the remote, peak height and inflection points of Se absorption using beamline 13B1 at the NSRRC, Taiwan. A good preliminary structure of XC1258 was obtained using the MAD approach with the *SOLVE* and *RESOLVE* programs (Hendrickson & Ogata, 1997; Terwilliger & Berendzen, 1999); the protein was found to adopt a novel quaternary structure variant of previously determined CN-hydrolase superfamily protein structures. Detailed structural refinement of XC1258 is currently under way.

A multiple sequence alignment of XC1258 with other CN-hydrolase superfamily proteins for which tertiary structures have been obtained (PDB codes 1f89, 1ems, 1erz and 1fo6) revealed very low overall sequence identities (23, 17, 14.2 and 16.1%, respectively) despite the presence of conserved residues in the catalytic triad motif regions (Fig. 4; Pace & Brenner, 2001). Of the three histidine residues (129, 144 and 215) identified to be important for catalytic activity in DCase (Wang *et al.*, 2001), only His129 was found to be conserved in XC1258; the other two residues His144 and His215 are replaced by Phe and Ala, respectively. Five gap-deletion regions were also observed in the XC1258 primary sequence compared with those of DCases (1fo6 and 1erz; Fig. 4). These sequence-alignment studies indicate that XC1258 possibly does not belong to the DCase subfamily of the CN superfamily. Further structural and functional studies are necessary to decipher its exact function.

This work is supported by an Academic Excellence Pursuit grant from the Ministry of Education and the National Science Council, Taiwan to S-HC. We thank the Core Facilities for Protein X-ray Crystallography in the Academia Sinica, Taiwan and the National Synchrotron Radiation Research Center, Taiwan for assistance during X-ray data collection. The National Synchrotron Radiation Research Center is a user facility supported by the National Science Council, Taiwan and the Protein Crystallography Facility is supported by the National Research Program for Genomic Medicine, Taiwan.

References

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. L. (2000). *Nucleic Acids Res.* **28**, 263–266.
Bork, P. & Koonin, E. V. (1994). *Protein Sci.* **3**, 1344–1346.

Corpet, F. (1988). *Nucleic Acids Res.* **16**, 10881–10890.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
Huang, W., Jia, J., Cummings, J., Nelson, M., Schneider, G. & Lindqvist, Y. (1997). *Structure*, **5**, 691–699.
Kumaran, D., Eswaramoorthy, S., Gerchman, S. E., Kycia, H., Studier, F. W. & Swaminathan, S. (2003). *Proteins*, **52**, 283–291.
Nagashima, S., Nakasako, M., Dohmae, N., Tsujimura, M., Takio, K., Odaka, M., Yohda, M., Kamiya, N. & Endo, I. (1998). *Nature Struct. Biol.* **5**, 347–351.
Nakai, T., Hasegawa, T., Yamashita, E., Yamamoto, M., Kumasaka, T., Ueki, T., Nanba, H., Ikenaka, Y., Takahashi, S., Sato, M. & Tsukihara, T. (2000). *Structure*, **8**, 729–739.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pace, H. C. & Brenner, C. (2001). *Genome Biol.* **2**, reviews0.1–0.9.
Pace, H. C., Hodawadekar, S. C., Draganescu, A., Huang, J., Bieganowski, P., Pekarsky, Y., Croce, C. M. & Brenner, C. (2000). *Curr. Biol.* **10**, 907–917.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Wang, W.-C., Hsu, W.-H., Chien, F.-T. & Chen, C.-Y. (2001). *J. Mol. Biol.* **306**, 251–261.
Wu, Y.-Y., Chin, K.-H., Chou, C.-C., Lee, C.-C., Shr, H.-L., Lyu, P.-C., Wang, A. H.-J. & Chou, S.-H. (2005). *Acta Cryst.* F**61**, 902–905.