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

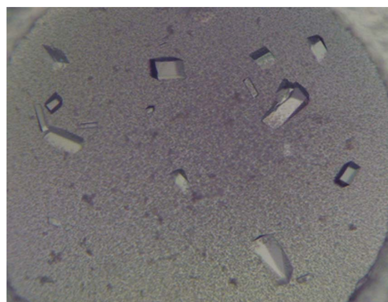
Purification, crystallization and preliminary crystallographic analysis of SMU.1108c protein from *Streptococcus mutans*

Ming-Jing Feng, Tian-Min Fu, Xiang Liu and Lan-Fen Li*

The National Laboratory of Protein Engineering and Plant Genetic Engineering, School of Life Sciences, Peking University, Beijing 100871, People's Republic of China

Correspondence e-mail: lilf@pku.edu.cn

Received 23 July 2010
Accepted 2 November 2010



© 2011 International Union of Crystallography
All rights reserved

Streptococcus mutans SMU.1108c (KEGG database) encodes a functionally uncharacterized protein consisting of 270 amino-acid residues. This protein is predicted to have a haloacid dehalogenase hydrolase-like domain and is a homologue of haloacid dehalogenase phosphatases that catalyze phosphoryl-transfer reactions. In this work, *SMU.1108c* was cloned into the pET28a vector and overexpressed in *Escherichia coli* strain BL21 (DE3). The protein was purified to homogeneity and crystallized using the sitting-drop vapour-diffusion method. The best crystal diffracted to 2.0 Å resolution and belonged to space group $C2$, with unit-cell parameters $a = 77.1$, $b = 80.2$, $c = 47.9$ Å, $\beta = 99.5°$.

1. Introduction

The haloacid dehalogenase (HAD) superfamily is one of the largest enzyme families, comprising a diverse set of over 6000 hydrolases that are encoded in the genomes of bacteria, archaea and eukaryotes (Kuznetsova *et al.*, 2006). The proteins of this superfamily are found to play roles in various cellular processes such as amino-acid biosynthesis and detoxification (Selengut, 2001; Koonin & Tatusov, 1994). Despite their low sequence identity (less than 15%; Roberts *et al.*, 2005), HAD-superfamily proteins share three conserved motifs and consequently a common catalytic mechanism (Koonin & Tatusov, 1994; Collet, Stroobant *et al.*, 1998).

Although the HAD superfamily is named after a dehalogenase, most members of the family have phosphoryl-transfer activity (Aravind & Koonin, 1998; Aravind *et al.*, 1998; Collet, van Schaftingen *et al.*, 1998). These HAD phosphatases/phosphohydrolases contain a conserved aspartic acid residue that serves as the nucleophile and consequently forms a phosphoaspartate intermediate (Collet, Stroobant *et al.*, 1998).

Streptococcus mutans, a Gram-positive bacterium, is one of the main aetiological agents of human dental caries, causing tooth decay and periodontal disease (Loesche, 1986; Ajdic *et al.*, 2002). There are 11 putative HAD proteins in *S. mutans*, including SMU.1108c (gi:24379541; Gene ID 1029554). The *SMU.1108c* gene encodes a putative hydrolase protein of 270 amino-acid residues with a theoretical molecular weight of 30.5 kDa. Protein-sequence analysis showed that SMU.1108c contains three conserved motifs and shares highest sequence similarity with HAD phosphatases (Altschul *et al.*, 1990). As part of the *S. mutans* structural genomics project which is in progress at Peking University (Su *et al.*, 2006), we report the purification, crystallization and preliminary X-ray analysis of SMU.1108c. Structural information on SMU.1108c might help us to identify its biological function and promote further studies of the HAD superfamily.

2. Materials and methods

2.1. Cloning

The *SMU.1108c* gene was obtained from the genomic DNA of *S. mutans* by polymerase chain reaction amplification (PCR; Saiki *et al.*, 1988). The primers used in the cloning were 5′-CGC**GGATCC**-ATGAGTGTGAAAGTTATTGCAA-3′ (SMU.1108c-F) and 5′-CCG-

CTCGAGCTAATCAATAGAAGCCAGATAA-3′ (SMU.1108c-R). The PCR product was inserted into the vector pET28a (Novagen) between the *Bam*HI and *Xho*I sticky-end restriction sites, creating a construct with an N-terminal His_6 tag (MGSSHHHHHHSSG-LVPRGSHMASMTGGQQMGRGS). The prepared recombinant plasmid was confirmed by DNA sequencing and transformed into the host *Escherichia coli* strain BL21 (DE3) (Invitrogen).

2.2. Protein expression and purification

The transformed cells were grown overnight at 310 K in 25 ml Luria–Bertani (LB) medium containing 50 µg ml^{-1} kanamycin. The overnight culture was inoculated into 1 l LB medium and grown at 310 K until the OD_{600} reached 0.6. The cells were then induced by adding isopropyl β-D-1-thiogalactopyranoside to a final concentration of 0.5 m*M* and grown for a further 5 h at 303 K. The cells were harvested by centrifugation at 6700*g* for 10 min and then resuspended in 20 ml buffer *A* (20 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl).

The suspension was disrupted by sonication on ice and the cell lysate was centrifuged at 35 000*g* for 40 min. After filtration, the supernatant was purified using an ÄKTA Explorer (GE Healthcare, Piscataway, New Jersey, USA) with the following steps. The supernatant was loaded onto a 5 ml Ni^{2+}-chelating affinity column (HiTrap, GE Healthcare, USA) previously equilibrated with buffer *A*. The column was washed using a linear gradient of buffer *B* (20 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl, 500 m*M* imidazole) from 0 to 100% in buffer *A* and the fractions containing the target protein were pooled and concentrated to about 1 ml at 277 K using a Millipore centrifugal ultrafiltration device (Amicon Ultra, 10 kDa cutoff). Further purification was carried out by size-exclusion chromatography on a HiLoad Superdex 75 column (column volume 120 ml; GE Healthcare) using buffer *C* (20 m*M* Tris–HCl pH 7.5, 200 m*M* NaCl). The fractions containing the target protein were pooled and concentrated to around 20 mg ml^{-1} for protein crystallization screening trials. The protein concentration was measured using the Bio-Rad Protein Assay Reagent (Bio-Rad Laboratories, USA) with bovine serum albumin (BSA) as the standard protein. After each step, the purified proteins were examined by SDS–PAGE (Fig. 1).

2.3. Crystallization

Crystallization was performed at 289 K by the sitting-drop vapour-diffusion method with an XtalQuest 48^2 crystallization plate (XtalQuest Inc., Beijing, People's Republic of China) using the commercial screening kits Crystal Screen, Crystal Screen 2, Natrix, Index (Hampton Research, California, USA) and BioXtal (XtalQuest Inc.) as initial screening conditions. 1 µl protein solution (20 mg ml^{-1} in buffer *C*) was mixed with 1 µl reservoir solution and equilibrated against 500 µl reservoir solution.

During the course of crystallization optimization, the purified fusion protein was digested using thrombin (Sigma; the specific cleavage site is LVPR↓GS) to remove the N-terminal His_6 tag. The digestion was performed at 277 K overnight by the incubation of one unit of thrombin with 1 mg protein. After thrombin digestion, the residues GSHMASMTGGQQMGRGS were left at the N-terminus of the protein. Without further purification, the digested protein was concentrated to ~20 mg ml^{-1} in buffer *C* for further crystallization screening.

2.4. Data collection

For data collection, a crystal was directly flash-cooled in liquid nitrogen after immersion in mother liquor without any additional cryoprotectant. X-ray diffraction data were collected on a Bruker SMART 6000 CCD using Cu $K\alpha$ radiation at a wavelength of 1.5418 Å from a Bruker MicrostarH rotating-anode generator operated at 45 kV and 40 mA. 1500 frames were collected with 0.3° oscillation per image at $2\theta = 0°$ and 1200 frames were collected with 0.3° oscillation per image at $2\theta = 10°$ with a crystal-to-detector distance of 59 mm at 100 K. The exposure time for each frame was 30 s. The two data sets were merged together. The diffraction data were processed using the *PROTEUM* software suite (Bruker AXS Inc., Madison, Wisconsin, USA).

3. Results and discussion

The SMU.1108c protein was overexpressed in a soluble form in *E. coli* BL21 (DE3) cells with an N-terminal His_6 tag. The typical yield was about 10 mg pure protein per litre of cell culture.

The protein was purified to homogeneity by two steps of purification (affinity and size-exclusion chromatography). During size-exclusion chromatography, pure protein was eluted with an estimated molecular mass of 35 kDa as determined from the elution volume. It matched the theoretical molecular weight of the SMU.1108c protein (30.5 kDa) plus ~4 kDa from the His_6 peptide tag remaining from the

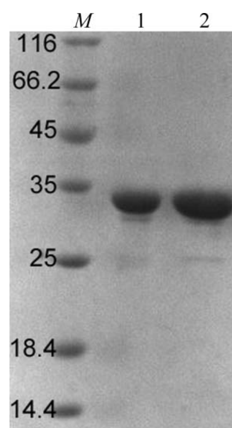


Figure 1
15% SDS–PAGE analysis of SMU.1108c protein purified by size-exclusion chromatography on a HiLoad Superdex 75 column (column volume 120 ml; GE Healthcare). Lane *M*, molecular-weight markers (kDa). Lanes 1 and 2, fractions of the target protein eluted with buffer *C* (20 m*M* Tris–HCl pH 7.5, 200 m*M* NaCl).

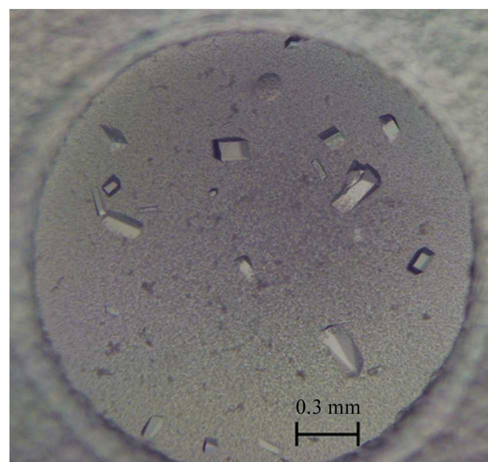


Figure 2
Crystals of SMU.1108c protein obtained with 0.1 *M* Tris–HCl pH 8.5, 25%(*w*/*v*) PEG 3350 using the sitting-drop vapour-diffusion method at 289 K.

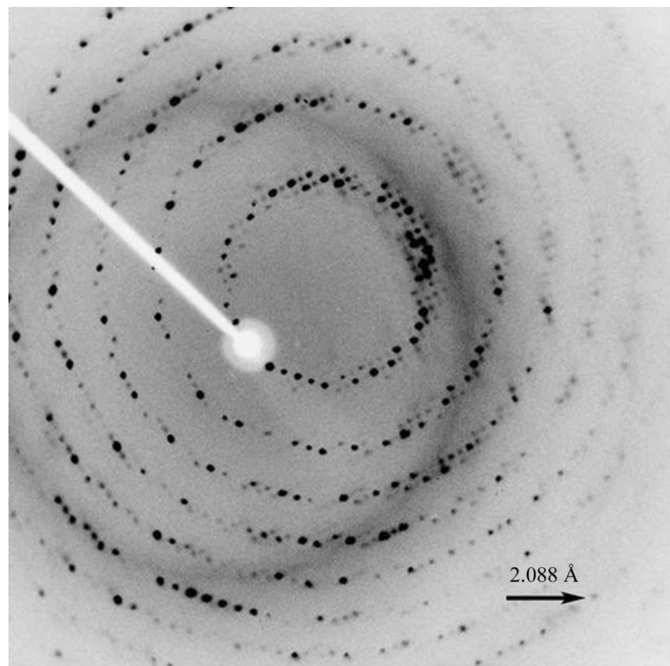


Figure 3
Diffraction image of the SMU.1108c protein crystal.

cloning procedure, indicating that the protein exists as a monomer in solution. Microcrystals were observed in 2–3 d under several conditions after initial protein crystallization screening. The crystals that were most suitable for diffraction experiments were obtained under the condition 0.1 *M* Tris–HCl pH 8.5 and 25%(*w*/*v*) PEG 3350 using fusion protein from which the N-terminal His_6 tag had been removed by thrombin digestion (Fig. 2). The average dimensions of the crystals were about 0.1 × 0.15 × 0.25 mm. The best crystal diffracted to a resolution of 2.0 Å (Fig. 3) and the data were finally processed to 2.15 Å as the overall completeness of the data was a little low (90.2%) when processing to high resolution (2.0 Å). The crystal belonged to space group *C*2, with unit-cell parameters $a = 77.1$, $b = 80.2$, $c = 47.9$ Å, $\beta = 99.5°$. With the assumption of one molecule per asymmetric unit, the calculated Matthews coefficient V_M (Matthews, 1968) was 2.4 $Å^3$ Da^{-1}, with a solvent content of 48.9%. The data-collection statistics are summarized in Table 1. The initial phase was obtained by molecular replacement using *MOLREP* (Vagin & Teplyakov, 2010) with *E. coli* YbiV (PDB code 1rlm; Roberts *et al.*, 2005) as the search model. Model building is being performed using the program *Coot* (Emsley & Cowtan, 2004). Meanwhile, an enzyme-activity assay is being carried out to explore the function of the protein in parallel with the structural studies. The three-dimensional structure will provide clues and evidence for the functional studies.

Table 1
Data-collection statistics for SMU.1108c.

Values in parentheses are for the highest resolution shell.

Resolution (Å)	54.95–2.15 (2.27–2.15)
Completeness (%)	97.8 (94.5)
R_{merge}† (%)	10.8 (9.5)
Average $I/\sigma(I)$	13.0 (7.7)
Space group	*C*2
Unit-cell parameters (Å, °)	$a = 77.1$, $b = 80.2$, $c = 47.9$, $\beta = 99.5$
No. of observed reflections	120347 (3538)
No. of unique reflections	15370 (2201)
Multiplicity	7.83 (1.6)
Mosaicity (°)	0.74
B factor ($Å^2$)	31.9
Molecules per asymmetric unit	1
Solvent content (%)	48.9

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

This work was supported by a grant from the National Natural Science Foundation of China (30530190). Peking University's 985 and 211 grants are also gratefully acknowledged.

References

Ajdic, D. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 14434–14439.
Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
Aravind, L., Galperin, M. Y. & Koonin, E. V. (1998). *Trends Biochem. Sci.* **23**, 127–129.
Aravind, L. & Koonin, E. V. (1998). *Trends Biochem. Sci.* **23**, 469–472.
Collet, J. F., Stroobant, V., Pirard, M., Delpierre, G. & Van Schaftingen, E. (1998). *J. Biol. Chem.* **273**, 14107–14112.
Collet, J. F., van Schaftingen, E. & Stroobant, V. (1998). *Trends Biochem. Sci.* **23**, 284.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Koonin, E. V. & Tatusov, R. L. (1994). *J. Mol. Biol.* **244**, 125–132.
Kuznetsova, E., Proudfoot, M., Gonzalez, C. F., Brown, G., Omelchenko, M. V., Borozan, I., Carmel, L., Wolf, Y. I., Mori, H., Savchenko, A. V., Arrowsmith, C. H., Koonin, E. V., Edwards, A. M. & Yakunin, A. F. (2006). *J. Biol. Chem.*, **281**, 36149–36161.
Loesche, W. J. (1986). *Microbiol. Rev.* **50**, 353–380.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 6491–6493.
Roberts, A., Lee, S.-Y., McCullagh, E., Silversmith, R. E. & Wemmer, D. E. (2005). *Proteins*, **58**, 790–801.
Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988). *Science*, **239**, 487–491.
Selengut, J. D. (2001). *Biochemistry*, **40**, 12704–12711.
Su, X.-D., Liang, Y., Li, L., Nan, J., Bröstromer, E., Liu, P., Dong, Y. & Xian, D. (2006). *Acta Cryst.* D**62**, 843–851.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.