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

# Protein preparation, crystallization and preliminary X-ray crystallographic analysis of SMU.961 protein from the caries pathogen *Streptococcus mutans*

**Xiong-Zhuo Gao,[a,b] Lan-Fen Li,[b] Xiao-Dong Su,[b] XiaoJun Zhao[a]* and Yu-He Liang[b]***

[a]Institute for Nanobiomedical Technology and Membrane Biology, West China Hospital, Sichuan University, Chengdu 610065, Sichuan, People's Republic of China, and [b]National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871, People's Republic of China

Correspondence e-mail: zhaoxj@scu.edu.cn, liangyh@pku.edu.cn

Received 18 July 2007
Accepted 4 September 2007

The *smu*.961 gene encodes a putative protein of 183 residues in *Streptococcus mutans*, a major pathogen in human dental caries. The gene was cloned into expression vector pET28a and expressed in a substantial quantity in *Escherichia coli* strain BL21 (DE3) with a His tag at its N-terminus. The recombinant protein SMU.961 was purified to homogeneity in a two-step procedure consisting of $Ni^{2+}$-chelating and size-exclusion chromatography. Crystals suitable for X-ray diffraction were obtained by the hanging-drop vapour-diffusion method and diffracted to 2.9 Å resolution at beamline I911-3, MAX-II-lab, Sweden. The crystal belonged to space group $C2$, with unit-cell parameters $a = 98.62$, $b = 73.73$, $c = 184.73$ Å, $\beta = 98.82^{\circ}$.

## 1. Introduction

*Streptococcus mutans*, a member of the viridans streptococci, is the aetiologic agent of dental caries (Loesche, 1986) and is also a causative agent of subacute infective endocarditis (Ullman *et al.*, 1988). The complete genome sequence of *S. mutans* UA159, a serotype c strain, was published in 2002. It contains 1963 ORFs (open reading frames), about 63% of which have been assigned putative functions (Ajdic *et al.*, 2002). SMU.961 from *S. mutans* is a protein for which no function has yet been assigned that contains 183 amino-acid residues and has a molecular weight of 19.91 kDa. Sequence comparison showed that most proteins with high sequence identity to SMU.961 are from the *Streptococcus* genus. Several proteins exhibiting high sequence identity to SMU.961 are annotated as macrophage infectivity potentiator (Mip) proteins. The sequence identity between SMU.961 and a putative Mip protein (gi:125718721) from *S. sanguinis* SK36 is 89%. In *Legionella pneumophila*, the Mip protein contributes to full virulence and intracellular survival (Cianciotto *et al.*, 1990; Wintermeyer *et al.*, 1995) and therefore might be a potential target for antibacterial drug design. Mip proteins exhibit peptidylprolyl *cis*/*trans*-isomerase (PPIase) activity and belong to the FK506-binding protein (FKBP) enzyme family (Fischer *et al.*, 1992), but there is no significant sequence homology between Mip proteins and FKBPs. There is also no conserved domain between Mip proteins and FKBPs. The conserved cysteines in some FKBPs (such as FKBP13 from *Arabidopsis thaliana*), which relate to redox regulation of the enzyme activity (Gopalan *et al.*, 2006), were not found in SMU.961 and Mip protein sequences. So far, it is still unclear whether SMU.961 has Mip or PPIase activity. The determination of the SMU.961 structure will help to identify the exact function of the protein and provide clues for further studies.

## 2. Experimental results

### 2.1. Protein expression and purification

To construct the expression plasmid, two primers containing *Bam*HI and *Xho*I restriction sites were designed as follows: SMU961-F, 5′-CGCGGATCCATGTCTAAATTTACTATACATACAA-3′, and SMU961-R, 5′-CCGCTCGAGCTATTTCACATACTGCTGTAACT-3′. The *smu.961* gene was amplified by polymerase chain reaction (PCR; Saiki *et al.*, 1988) from *S. mutans* genomic DNA. The

© 2007 International Union of Crystallography
All rights reserved

PCR-amplified fragment was then cloned into the expression vector pET28a by conventional cloning methods with an N-terminal $His_6$ tag (MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGS). The correct insertion was confirmed by DNA sequencing; the expressed part of the constructed plasmid pET28a-SMU.961 contains 34 residues of N-terminal $His_6$ tag followed by 183 residues of SMU.961 protein. The constructed plasmid was transformed into *Escherichia coli* strain BL21 (DE3) cells for expression. The BL21 (DE3) cells containing pET28a-SMU.961 were grown overnight in 20 ml Luria–Bertani (LB) medium containing 50 µg ml$^{-1}$ kanamycin at 310 K; the culture was then added into 1 l LB medium containing 50 µg ml$^{-1}$ kanamycin. When the $OD_{600}$ reached 0.6, the culture was induced with isopropyl $\beta$-D-1-thiogalactopyranoside at a final concentration of 0.5 m*M* and expression was continued for 20 h at 291 K. The cells were harvested by centrifugation (6500 rev min$^{-1}$, 10 min, 277 K), resuspended in 20 ml buffer *A* (20 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl) and lysed by sonication on ice. After centrifugation at 277 K and 17 000 rev min$^{-1}$ for 60 min, the supernatant was collected and loaded onto a 5 ml $Ni^{2+}$-chelating affinity column (HiTrap, GE Healthcare, USA) previously equilibrated with buffer *A*. Impurities were washed out with 10% buffer *B* (20 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl, 500 m*M* imidazole) and the target protein was eluted with a linear gradient of 10–100% buffer *B*. The eluted protein was further purified by size-exclusion chromatography (HiLoad Superdex 75 XK16/60, GE Healthcare, USA) using buffer *C* (20 m*M* Tris–HCl pH 7.5, 200 m*M* NaCl). It was deduced from gel-filtration chromatography that the protein exists as a monomer in solution. The purity of the SMU.961 protein was examined by SDS–PAGE during each step. This showed that the purified protein has a molecular weight of about 24 kDa, in agreement with the predicted molecular weight (with an additional 4 kDa fusion part).

### 2.2. Protein crystallization

The purified protein from the gel filtration (in 20 m*M* Tris–HCl pH 7.5, 200 m*M* NaCl) was concentrated to 18 mg ml$^{-1}$ by ultrafiltration (Millipore, Amicon). Initial crystallization screening was carried out with Hampton Index, Crystal Screen, Crystal Screen 2 and Natrix kits (Hampton Research, CA, USA) using the hanging-drop vapour-diffusion method at 289 K. 1 µl protein solution was mixed with 1 µl reservoir solution and equilibrated against 500 µl reservoir solution.

Microcrystals were obtained in 3–4 d under several conditions. After subsequent optimization starting from condition No. 26 of Natrix, crystals suitable for X-ray diffraction were obtained from a condition containing 18 mg ml$^{-1}$ protein, 0.2 *M* potassium chloride, 0.1 *M* magnesium acetate, 0.05 *M* sodium cacodylate pH 5.9 and 10%(*w*/*v*) PEG 8000 (Fig. 1).

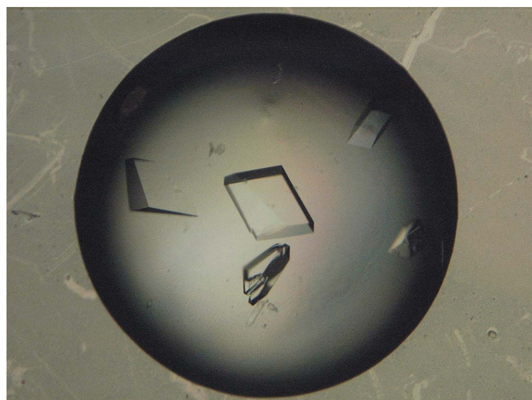

**Figure 1**
Crystals of SMU.961 protein.

**Table 1**
Data-collection statistics for the SMU.961 crystal.

Values in parentheses are for the highest resolution shell.

| | |
|---|---|
| Resolution (Å) | 50–2.9 (3.07–2.9) |
| Completeness (%) | 96.2 (89.0) |
| $R_{sym}$† (%) | 8.4 (27.0) |
| Mean $I/\sigma(I)$ | 10.5 (4.1) |
| Space group | *C*2 |
| Unit-cell parameters (Å, °) | $a$ = 98.62, $b$ = 73.73, $c$ = 184.73, $\beta$ = 98.82 |
| No. of observed reflections | 90324 |
| No. of unique reflections | 2844 |
| Molecules per ASU | 4–8 |

† $R_{sym} = \sum_h [\sum_i |I(h)_i - \langle I(h) \rangle| / \sum_i I(h)_i]$, where $I(h)_i$ is the $i$th observation of reflection $h$ and $\langle I(h) \rangle$ is the mean intensity of all observations of reflection $h$.

### 2.3. X-ray diffraction data collection and processing

The diffraction data were collected on a MAR Mosaic 225 CCD detector at beamline I911-3, MAX-II-lab, Lund, Sweden. The crystal was flash-cooled without cryoprotectant and maintained at 100 K in a cold nitrogen stream during data collection. The crystal-to-detector distance was set to 330 mm and the wavelength was 0.979 Å. A total of 160 frames were collected with 1° $\varphi$ oscillation per frame. Data were processed using the *XDS* program (Kabsch, 1993). The merged data set has a completeness of 96.2% to 2.9 Å resolution and has an overall $R_{sym}$ of 8.4%. The space group was determined to be *C*2, with unit-cell parameters $a$ = 98.62, $b$ = 73.73, $c$ = 184.73 Å, $\beta$ = 98.82°. There could be four to eight molecules per asymmetric unit, which corresponds to a solvent content of between 29.4 and 64.7% (Matthews, 1968). Statistics for diffraction data collection and processing are summarized in Table 1. Since there is no homologous model in the PDB which could serve as a search model for molecular-replacement calculations, SeMet protein was produced and phase determination using anomalous dispersion is ongoing.

This work was supported by a grant from the National Natural Science Foundation of China (30530190). Sichuan University's 985 grants are greatly acknowledged. Peking University's 985 and 211 grants are also greatly acknowledged. Y-HL is the recipient of a Fok Ying Tong Education Foundation grant (94017). We also thank the beamline staff at MAX-II-lab, Lund, Sweden for their kind help with data collection.

## References

Ajdic, D. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 14434–14439.
Cianciotto, N. P., Eisenstein, B. I., Mody, C. H. & Engleberg, N. C. (1990). *J. Infect. Dis.* **162**, 121–126.
Fischer, G., Bang, H., Ludwig, B., Mann, K. & Hacker, J. (1992). *Mol. Microbiol.* **6**, 1375–1383.
Gopalan, G., He, Z., Battaile, K. P., Luan, S. & Swaminathan, K. (2006). *Proteins*, **65**, 789–795.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Loesche, W. J. (1986). *Microbiol. Rev.* **50**, 353–380.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
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.
Ullman, R. F., Miller, S. J., Strampfer, M. J. & Cunha, B. A. (1988). *Heart Lung*, **17**, 209–212.
Wintermeyer, E., Ludwig, B., Steinert, M., Schmidt, B., Fischer, G. & Hacker, J. (1995). *Infect. Immun.* **63**, 4576–4583.