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

# Preliminary X-ray crystallographic analysis of SMU.2055 protein from the caries pathogen *Streptococcus mutans*

**Wang-Hong Zhao,[a,b] Xiu-Rong Zhan,[b] Xiong-Zhuo Gao,[c] Xiang Liu,[c] Yi-Fei Zhang,[a] Jiuxiang Lin,[a] Lan-Fen Li,[c]* Shi-Cheng Wei[c]* and Xio-Dong Su[c]**

[a]School of Stomatology, Peking University, Beijing 100081, People's Republic of China, [b]School of Stomatology, Lanzhou University, Lanzhou 730000, People's Republic of China, and [c]The 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: lilf@pku.edu.cn, kqsc-wei@bjmu.edu.cn

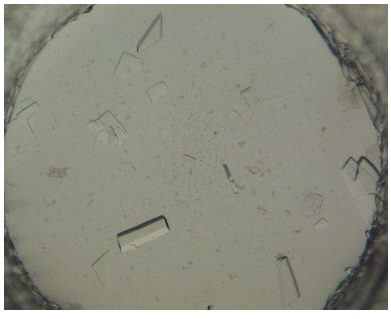
Received 25 December 2009
Accepted 19 March 2010

The SMU.2055 gene from the major caries pathogen *Streptococcus mutans* is annotated as a putative acetyltransferase with 163 amino-acid residues. In order to identify its function *via* structural studies, the SMU.2055 gene was cloned into the expression vector pET28a. Native and SeMet-labelled SMU.2055 proteins with a $His_6$ tag at the N-terminus were expressed at a high level in *Escherichia coli* strain BL21 (DE3) and purified to homogeneity by $Ni^{2+}$-chelating affinity chromatography. Diffraction-quality crystals of SeMet-labelled SMU.2055 were obtained using the sitting-drop vapour-diffusion method and diffracted to a resolution of 2.5 Å on beamline BL17A at the Photon Factory, Tsukuba, Japan. The crystals belong to the orthorhombic space group $C222_1$, with unit-cell parameters $a$ = 92.0, $b$ = 95.0, $c$ = 192.2 Å. The asymmetric unit contained four molecules, with a solvent content of 57.1%.

## 1. Introduction

Dental caries is one of the most prevalent diseases afflicting humans. It is an infectious disease that not only affects the calcified tissue of teeth (Shivakumar *et al.*, 2009) but can also lead to a series of diseases such as pulpitis (Nguyen & Martin, 2008) and periapical abscesses (Mueller & Lowder, 1998). The marked increase in the prevalence of dental caries signals a pending public health crisis (Bagramian *et al.*, 2009). *Streptococcus mutans* has been recognized as the leading causative agent of human dental caries (Loesche, 1986) and recent studies on the cariogenesis mechanism have focused on biofilm formation (Bleiweis *et al.*, 1992), acid tolerance (Quivey *et al.*, 2001), virulence factors and bacterial adhesion to the tooth (Islam *et al.*, 2007). In addition, *S. mutans* can cause subacute bacterial endocarditis (Ullman *et al.*, 1988).

The genome of *S. mutans* UA159, a serotype C strain, has been sequenced and contains about 1960 open reading frames, 63% of which have been assigned putative functions (Ajdic *et al.*, 2002). An *S. mutans* structural genomics project aimed at solving the structures of the majority of the cytosolic proteins was initiated in the People's Republic of China in 2005 (Su *et al.*, 2006). One of the selected targets, SMU.2055 (gi:24380392; Gene ID 1029233), has successfully been crystallized. The SMU.2055 gene from the *S. mutans* genome encodes a putative acetyltransferase protein with 163 amino-acid residues, a theoretical molecular weight of 18.6 kDa and an isoelectric point of 7.7. A *BLAST* search showed that SMU.2055 belongs to the Gcn5-related *N*-acetyltransferase (GNAT) superfamily. The GNAT superfamily is widespread in nature and contains numerous members that use acyl-CoAs to acylate their cognate substrates (Vetting *et al.*, 2005). SMU.2055 and its closest homologue in the Protein Data Bank (PDB), a GNAT-family acetyltransferase from *Enterococcus faecalis* V583 (PDB code 2ae6; Y. Kim, C. Hatzos, S. Moy, F. Collart & A. Joachimiak, unpublished work), share only 27% sequence identity. As it may be difficult to solve the SMU.2055 structure by molecular



replacement with such a low sequence identity, we therefore prepared selenomethionine-labelled SMU.2055 for use in phase determination (Doublié, 1997). Structure determination of SMU.2055 will help us to better study its biological function.

## 2. Experimental procedures

### 2.1. Gene cloning

To construct the expression plasmid, two primers containing restriction sites were designed: SMU.2055-F (5′-CGCGGATCCATGAAAATAAGCCCTATGTTA-3′) and SMU.2055-R (5′-CCGCTCGAGTTATTTGGCATAGGCAGCCT-3′). The SMU.2055 gene was amplified by polymerase chain reaction (PCR) from *S. mutans* genomic DNA (Saiki *et al.*, 1988). After digestion with *Bam*HI and *Xho*I overnight at room temperature, the PCR product was cloned into the vector pET28a(+)(Novagen) by the conventional cloning method with an N-terminal fusion $His_6$ tag (MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGS). The recombinant vector containing the target gene SMU.2055, which was verified by DNA sequencing, was transformed into *Escherichia coli* strain BL21 (DE3) for protein expression.

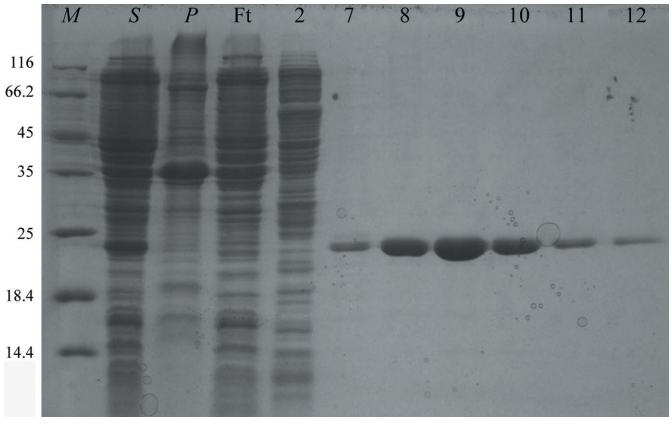



**Figure 1**
15% SDS–PAGE analysis of SMU.2055 protein purified by $Ni^{2+}$-affinity column chromatography (with Coomassie Brilliant Blue staining). Lane *M*, molecular-weight markers (kDa). Lanes *S* and *P*, soluble and insoluble material after inducing expression of SMU.2055, respectively. Lane Ft, unbound material from the $Ni^{2+}$-affinity column; lane 2, fraction eluted with 10% elution buffer (20 m*M* Tris–HCl pH 8.5, 500 m*M* NaCl and 500 m*M* imidazole) in lysis buffer (20 m*M* Tris–HCl pH 8.5, 500 m*M* NaCl). Lanes 7–12, fractions of the target protein eluted with a linear gradient of elution buffer from 10 to 100% in lysis buffer.

### 2.2. Protein expression and purification

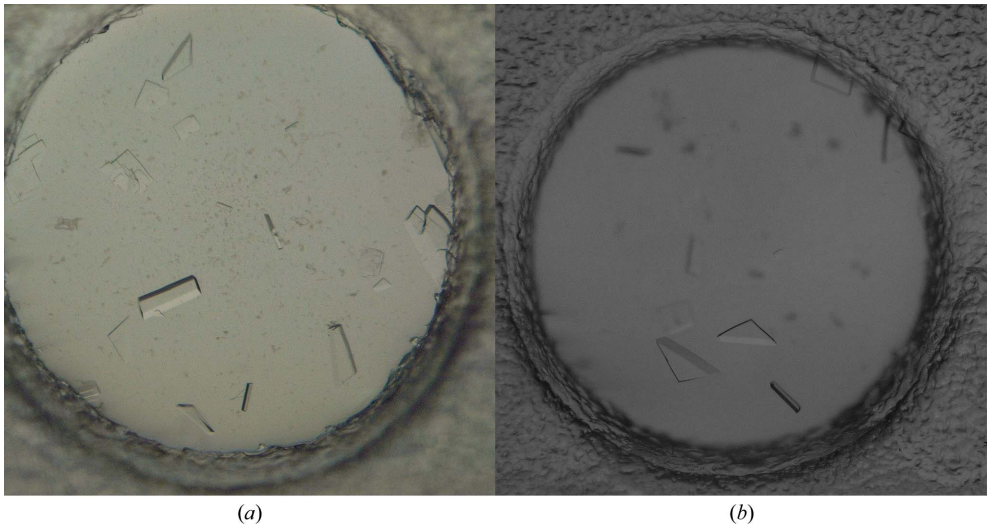
*E. coli* strain BL21 (DE3) cells containing pET28a-SMU.2055 were grown overnight in 20 ml Luria–Bertani (LB) medium containing 50 µg ml$^{-1}$ kanamycin at 310 K. The overnight culture was transferred into 1 l LB medium containing 50 µg ml$^{-1}$ kanamycin and growth continued at 310 K until the $OD_{600}$ reached 0.6. The cells were then induced with 0.5 m*M* isopropyl *β*-D-1-thiogalactopyranoside (IPTG) for 8 h at 303 K.

The cells were harvested by centrifugation for 10 min at 8600*g* and 277 K, resuspended in 20 ml lysis buffer (20 m*M* Tris–HCl, 500 m*M* NaCl pH 8.5) and lysed by sonication on ice. After centrifugation at 48 400*g* and 277 K for 20 min twice to remove debris, the supernatant was filtered and loaded onto a 5 ml HiTrap $Ni^{2+}$-chelating affinity column (GE Healthcare, USA) equilibrated with lysis buffer. The unbound material was first washed with lysis buffer. The low-nickel-affinity proteins were eluted with 10% elution buffer (20 m*M* Tris–HCl pH 8.5, 500 m*M* NaCl and 500 m*M* imidazole) in lysis buffer. Subsequently, the bound target protein was eluted with a 10–100% linear gradient of elution buffer in lysis buffer and examined by SDS–PAGE (Fig. 1). The fractions containing the target protein were pooled and concentrated by ultrafiltration using a Millipore centrifugal ultrafiltration device (Amicon Ultra, 10 kDa cutoff) at 277 K. The target protein was buffer-exchanged into buffer solution (20 m*M* Tris–HCl pH 8.5, 500 m*M* NaCl) by ultrafiltration.

At the same time, SeMet-labelled SMU.2055, which was used for phase determination, was prepared and purified using a similar procedure.

### 2.3. Protein crystallization

For crystallization trials, the purified native and SeMet-labelled SMU.2055 proteins were concentrated to 19 and 27 mg ml$^{-1}$, respectively, in buffer solution (20 m*M* Tris–HCl pH 8.5, 500 m*M* NaCl) by ultrafiltration without removal of the N-terminal fusion



(*a*) (*b*)

**Figure 2**
Crystals of (*a*) native and (*b*) SeMet-labelled SMU.2055 protein. Both were obtained using a reservoir solution consisting of 30%(*v*/*v*) pentaerythritol ethoxylate (15/4 EO/OH), 0.05 *M* bis-tris pH 6.5 and 0.05 *M* ammonium sulfate.

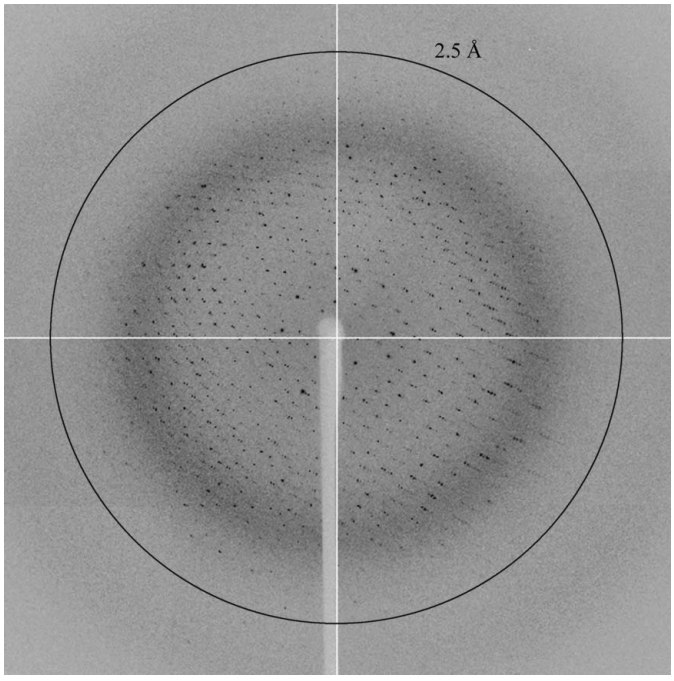



**Figure 3**
Diffraction image from a SeMet-labelled SMU.2055 protein crystal.

$His_6$ tag. The protein concentration was determined using a Bio-Rad protein-assay kit (Bio-Rad Laboratories, USA) based on the method of Bradford. Initial crystallization screening was carried out by the sitting-drop vapour-diffusion method with an XtalQuest48[2] crystallization plate (XtalQuest Inc., Beijing, People's Republic of China) at 289 K, using several commercially available crystallization screening kits: Index, Crystal Screen, Crystal Screen 2 and Natrix (Hampton Research, USA) and BioXtal (XtalQuest Inc., Beijing, People's Repubic of China). 1 µl protein solution was mixed with an equal volume of reservoir solution and equilibrated against 130 µl reservoir solution.

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

Native and SeMet-labelled SMU.2055 protein crystals were quick-soaked (~30 s) in reservoir solution containing 25%(*v*/*v*) glycerol as a cryoprotectant and flash-cooled in liquid nitrogen. The native crystal diffracted to a maximum resolution of 2.3 Å using a Bruker SMART 6000 CCD and Cu $K\alpha$ radiation from a Bruker MICROSTAR-H rotating-anode generator operated at 45 kV and 40 mA. 800 frames were collected with 0.3° oscillation per image. The diffraction data were processed using the *PROTEUM* software suite (Bruker). Diffraction data from the SeMet-labelled SMU.2055 protein crystal were collected on beamline BL17A at the Photon Factory (Tsukuba, Japan) at the peak wavelength of 0.97909 Å. The crystal-to-detector distance was set to 276.4 mm. 360 frames were collected with 1° oscillation per image. The diffraction data were processed using the *XDS* program suite (Kabsch, 1993).

## 3. Results

The typical yield of the native protein was about 46 mg pure protein per litre of cell culture; the SeMet-labelled protein showed a similar yield. The SDS–PAGE results showed that the purified protein had a molecular mass of about 22.6 kDa, which is consistent with the calculated molecular mass of the SMU.2055 protein of 18.6 kDa plus a 4 kDa $His_6$ tag.

**Table 1**
Data-collection statistics for native and SeMet-labelled SMU.2055 protein crystals.

Values in parentheses are for the highest resolution shell.

| | Native | SeMet-labelled |
|---|---|---|
| Wavelength (Å) | 1.54 | 0.97909 |
| Resolution (Å) | 50–2.3 (2.4–2.3) | 50.0–2.5 (2.6–2.5) |
| Completeness (%) | 82.2 (38.9) | 99.4 (97.0) |
| $R_{merge}$† (%) | 7.7 (36.7) | 9.1 (49.2) |
| Mean $I/\sigma(I)$ | 8.9 (1.8) | 18.8 (4.4) |
| Space group | $C222_1$ | $C222_1$ |
| Unit-cell parameters (Å) | $a$ = 92.0, $b$ = 94.6, $c$ = 193.9 | $a$ = 92.0, $b$ = 95.0, $c$ = 192.2 |
| No. of observed reflections | 223053 (14746) | 405912 (32609) |
| No. of unique reflections | 31240 (1739) | 56047(6048) |
| No. of molecules in the asymmetric unit | 4 | 4 |
| $V_M$ (Å$^3$ Da$^{-1}$) | 2.84 | 2.86 |
| Solvent content (%) | 56.7 | 57.1 |

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of reflection $hkl$ and $\sum_i$ is the sum over all $i$ measurements of reflection $hkl$.

Crystals of the native and SeMet-labelled protein suitable for diffraction were obtained using the condition 30%(*v*/*v*) pentaerythritol ethoxylate (15/4 EO/OH), 0.05 *M* bis-tris pH 6.5 and 0.05 *M* ammonium sulfate (BioXtal screen condition No. 95). The crystal dimensions of the native SMU.2055 protein crystals were about 0.1 × 0.1 × 0.5 mm (Fig. 2*a*); the SeMet-labelled SMU.2055 protein crystals were not easily obtained but grew to dimensions of about 0.1 × 0.2 × 0.3 mm (Fig. 2*b*).

The SeMet-labelled SMU.2055 protein crystals diffracted to a resolution of 2.5 Å and belonged to space group $C222_1$ as indicated by systematic absences, with unit-cell parameters $a$ = 92.0, $b$ = 95.0, $c$ = 192.2 Å. Assuming the presence of four molecules per asymmetric unit, the $V_M$ value is 2.86 Å$^3$ Da$^{-1}$, corresponding to a solvent content of 57.1% (Matthews, 1968). Fig. 3 shows a typical diffraction pattern for the SeMet-labelled crystals. The native crystal belonged to the same space group, with very similar unit-cell parameters. The crystallographic parameters and data-collection statistics for both the native and SeMet-labelled proteins are listed in Table 1. The phases of the structure have been determined using the single-wavelength anomalous dispersion method and refinement of the structure is in progress.

This work was supported by a grant from the National Natural Science Foundation of China (30530190). Peking University's 985 and 211 grants are greatly acknowledged. We are also grateful for the support of the Photon Factory, KEK, Tsukuba, Japan during data collection.

## References

Ajdic, D. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 14434–14439.
Bagramian, R. A., Garcia-Godoy, F. & Volpe, A. R. (2009). *Am. J. Dent.* **22**, 3–8.
Bleiweis, A. S., Oyston, P. C. & Brady, L. J. (1992). *Adv. Exp. Med. Biol.* **327**, 229–241.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Islam, B., Khan, S. N. & Khan, A. U. (2007). *Med. Sci. Monit.* **13**, RA196–RA203.
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.
Mueller, P. O. & Lowder, M. Q. (1998). *Vet. Clin. North Am. Equine Pract.* **14**, 349–363.
Nguyen, D. H. & Martin, J. T. (2008). *Am. Fam. Physician*, **77**, 797–802.

Quivey, R. G., Kuhnert, W. L. & Hahn, K. (2001). *Crit. Rev. Oral Biol. Med.* **12**, 301–314.
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.
Shivakumar, K. M., Vidya, S. K. & Chandu, G. N. (2009). *Indian J. Dent. Res.* **20**, 99–106.
Su, X.-D., Liang, Y., Li, L., Nan, J., Bröstromer, E., Liu, P., Dong, Y. & Xian, D. (2006). *Acta Cryst.* D**62**, 843–851.
Ullman, R. F., Miller, S. J., Strampfer, M. J. & Cunha, B. A. (1988). *Heart Lung*, **17**, 209–212.
Vetting, M. W., de Carvalho, L. P. S., Yu, M., Hegde, S. S., Magnet, S., Roderick, S. L. & Blanchard, J. S. (2005). *Arch. Biochem. Biophys.* **433**, 212–226.