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

Preliminary X-ray crystallographic analysis of SMU.573, a putative sugar kinase from *Streptococcus mutans*

Yan-Feng Zhou,[a] Lan-Fen Li,[a] Cheng Yang,[a,b] Yu-He Liang[a]* and Xiao-Dong Su[a,c]*

[a]National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871, People's Republic of China, [b]Rigaku/MSC Inc., 9009 New Trails Drive, The Woodlands, TX 77381, USA, and [c]Shenzhen Graduate School of Peking University, Shenzhen 518055, People's Republic of China

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

Received 8 October 2007
Accepted 5 December 2007

SMU.573 from *Streptococcus mutans* is a structurally and functionally uncharacterized protein that was selected for structural biology studies. Native and SeMet-labelled proteins were expressed with an N-His tag in *Escherichia coli* BL21 (DE3) and purified by Ni^{2+}-chelating and size-exclusion chromatography. Crystals of the SeMet-labelled protein were obtained by the hanging-drop vapour-diffusion method and a 2.5 Å resolution diffraction data set was collected using an in-house chromium radiation source. The crystals belong to space group $I4$, with unit-cell parameters $a = b = 96.53$, $c = 56.26$ Å, $\alpha = \beta = \gamma = 90^\circ$.

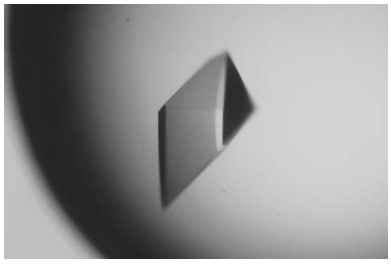
1. Introduction

SMU.573 from *Streptococcus mutans* is a structurally and functionally uncharacterized protein consisting of 277 amino-acid residues with a molecular weight of 30 kDa. A *BLAST* search showed that SMU.573 homologues are widely distributed (Fig. 1) among bacteria. Most of the homologues are annotated as 'function unknown', but a few of them are predicted to be sugar kinases, carbohydrate kinases or ribokinase-like proteins. There are no structures of highly similar homologues to SMU.573 in the PDB. The proteins in the PDB that show some sequence identity to SMU.573 are YXKO from *Bacillus subtilis* (PDB code 1kyh; 31% sequence identity; Zhang *et al.*, 2002) and Tm0922 from *Thermotoga maritima* (PDB code 2ax3; 28% sequence identity). These two proteins show structural similarity to ribokinase-like carbohydrate kinase superfamily proteins, which fold into a three-layer α/β sandwich. The members of the ribokinase-like superfamily are all phosphotransferases and catalyze the phosphorylation of the hydroxymethyl group of the substrate; magnesium and ATP (or ADP) are required for their activity (Sigrell *et al.*, 1998; Campobasso *et al.*, 2000; Cheng *et al.*, 2002). Ribokinase-like superfamily members play important roles in ribose utilization (Barsotti & Ipata, 2002; Campobasso *et al.*, 2000) and could be potential targets for structure-based drug design. The structural determination of SMU.573 would provide clues for functional studies of this protein family and could be helpful in antimicrobial design against *S. mutans*, the leading pathogen of human dental caries (Loesche, 1986).

2. Experimental procedures and results

2.1. Gene cloning and protein expression

The SMU.573 gene was amplified by the polymerase chain reaction (PCR) from *S. mutans* genomic DNA. The amplified fragment was inserted into the *Bam*HI/*Xho*I-digested expression vector pET-28a(+) (Novagen). An N-terminal tag of 34 amino-acid residues including six histidines was added to the gene product (MGSSH-HHHHHSSGLVPRGSHMASMTGGQQMGRGS) and the correct insertion was confirmed by DNA sequencing. The plasmid was transformed into *Escherichia coli* BL21 (DE3) cells for expression. The transformed cells were cultured in LB medium supplemented with 50 µg ml^{-1} kanamycin at 310 K until the OD_{600} reached 0.6–0.8. Protein expression was then induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to the culture medium to a final



concentration of 0.5 m*M* and the culture was incubated for a further 20 h at 291 K. The cells were then harvested by centrifugation at 6700*g* for 10 min, resuspended in lysis buffer containing 20 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl and disrupted by sonication on ice. The lysate was centrifuged at 34 700*g* at 277 K for 30 min twice to exclude the debris. The supernatant was loaded onto a 5 ml HiTrap Ni column (GE Healthcare, USA) equilibrated with buffer containing 20 m*M* Tris–HCl pH 7.5, 0.5 *M* NaCl. The unbound proteins were eluted with the equilibration buffer and the loosely bound proteins were eluted with buffer containing 20 m*M* Tris–HCl pH 7.5, 0.5 *M* NaCl, 75 m*M* imidazole. The tightly bound proteins were then eluted with buffer containing 20 m*M* Tris–HCl pH 7.5, 0.5 *M* NaCl and 500 m*M* imidazole. The protein was further purified on a HiLoad Superdex 75 column (GE Healthcare, USA) equilibrated with 20 m*M* Tris–HCl pH 7.5, 200 m*M* NaCl. The protein eluted as a dimer. The purity of the protein was examined by SDS–PAGE at each step. The monomer has a molecular weight of about 34 kDa, in agreement with the predicted molecular weight and including the 4 kDa tag. To accelerate the structure determination, selenomethionine-labelled (SeMet-labelled) SMU.573 was prepared at the same time as the native protein using the method described by Doublié (1997). The purification procedure of the SeMet-labelled protein was identical to that of the wild-type protein. The yields were about 100 mg per litre for both native and SeMet-labelled proteins.

2.2. Crystallization, X-ray data collection and processing

The purified wild-type and SeMet-labelled SMU.573 proteins were concentrated to 20 mg ml^{-1} in the the gel-filtration buffer by ultrafiltration (Ultra-15, 10 kDa cutoff, Millipore Amicon) and used for crystallization experiments. Initial crystallization screening was carried out using the Index, Crystal Screen and Crystal Screen 2 screening kits (Hampton Research, California, USA) and the hanging-drop vapour-diffusion method at 289 K. 1 µl protein solution at 20 mg ml^{-1} was mixed with 1 µl reservoir solution and the droplets were equilibrated against 500 µl reservoir solution. Crystals of

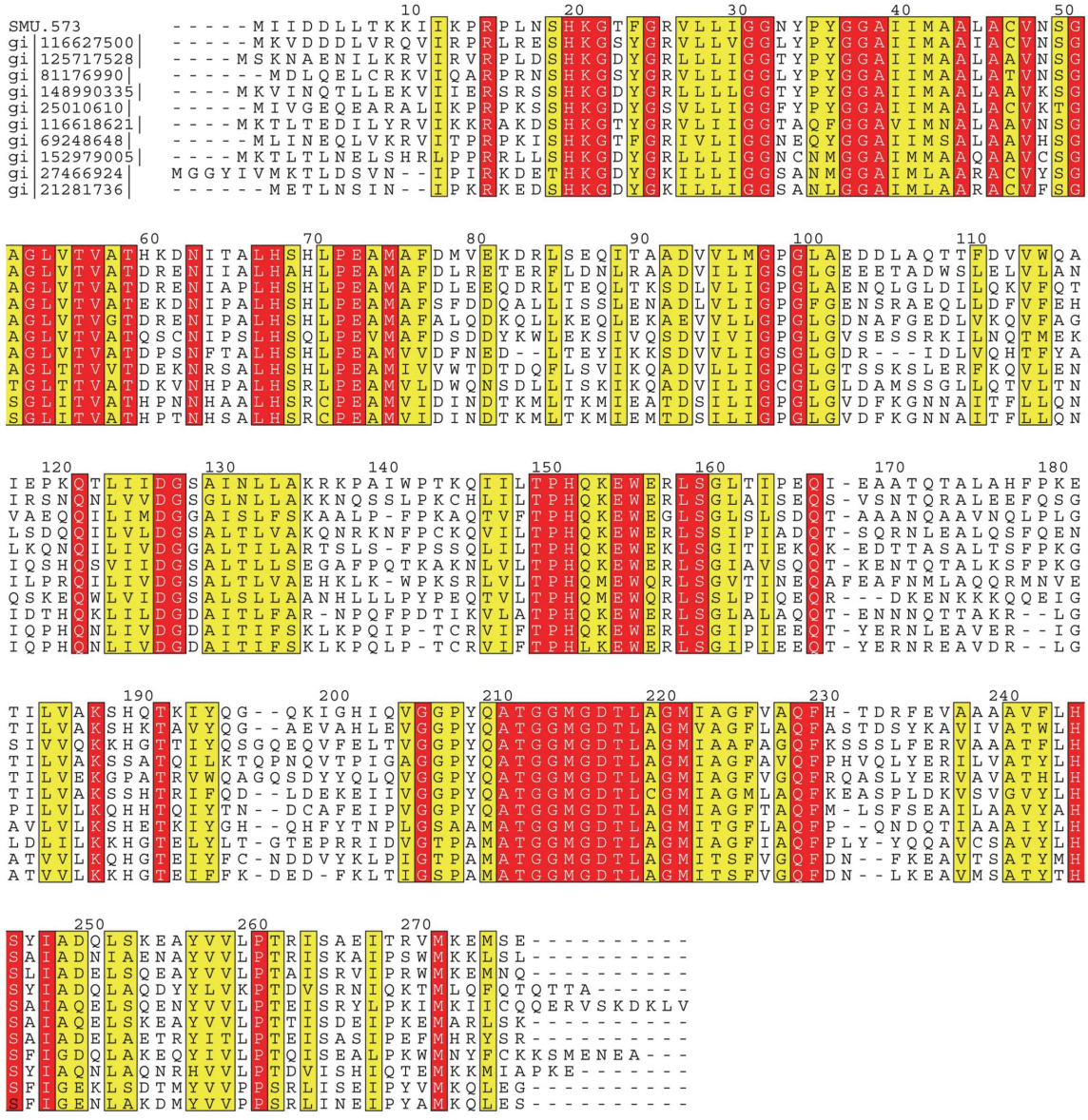


Figure 1
Sequence alignment of SMU.573 homologues. The alignment was performed using the programs *ClustalX* (Thompson *et al.*, 1997) and *ALSCRIPT* (Barton, 1993); strictly and highly conserved residues are marked with red and yellow backgrounds, respectively.

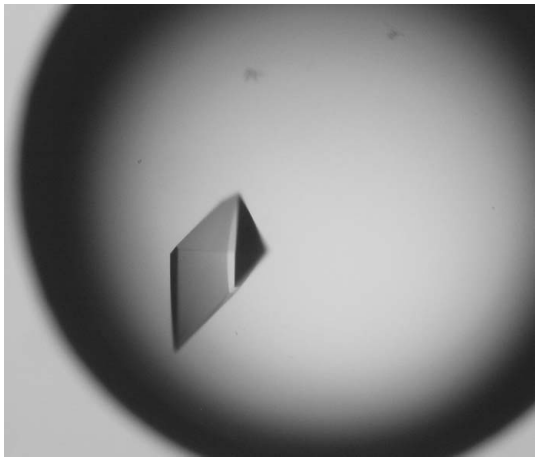


Figure 2
A crystal of SeMet-labelled SMU.573 protein. Approximate dimensions are 0.1 × 0.1 × 0.3 mm.

SeMet-labelled SMU.573 suitable for diffraction experiments were obtained after 4–5 d with 0.49 *M* $NaH_2PO_4.H_2O$ and 0.91 *M* K_2HPO_4 pH 6.9 (Fig. 2). The dimensions of the crystals were about 0.1 × 0.1 × 0.3 mm. Crystals of the native protein were also obtained but did not diffract very well and are still in the process of optimization.

Solid sucrose particles were added to the mother liquor as a cryoprotectant. A single crystal of SeMet-labelled SMU.573 was transferred to a drop of mother liquor containing the cryoprotectant for about 10 s and then flash-cooled in liquid nitrogen. The crystal was mounted in an arbitrary orientation and data were collected at cryogenic temperature in a single continuous scan on a Rigaku MicroFocus-007HF with Cr VariMax multilayer optics and an R-AXIS IV^{++} image-plate detector. Since absorption can be more significant with Cr $K\alpha$ radiation, the loop size was selected to be slightly larger than the crystal in order to reduce the effects from the loop material and solvent. An exposure time of 60 s per degree was used for data collection in order to minimize radiation damage to the crystals. A total of 360° of data were collected. In order to reduce the loss of reflection signals, a pyramid-shaped helium path was installed in the front of the R-AXIS IV^{++} detector. The data were processed and integrated with *HKL*-2000 (Otwinowski & Minor, 1997). The crystal diffracted to better than 2.3 Å, but data were processed to 2.5 Å owing to low completeness in the outer shell. The unit-cell parameters for the crystal were $a = b = 96.53$, $c = 56.26$ Å, $\alpha = \beta = \gamma = 90°$ and the space group was assigned as *I*4. The systematic absences along the *c* axis were $l = 2n + 1$ (where *n* is an integer). Assuming the presence of one monomer in the asymmetric unit, the Matthews coefficient is 1.93 $Å^3\ Da^{-1}$, which gives a solvent content of 36.2% (Matthews, 1968). Details of the data-collection statistics are listed in Table 1. Structure determination by the single-wavelength anomalous dispersion method is ongoing.

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	2.2897
Resolution (Å)	50–2.5 (2.59–2.5)
Completeness (%)	94.0 (69.7)
R_{merge}† (%)	5.3 (7.8)
Mean $I/\sigma(I)$	29.5 (20.4)
Multiplicity	10.9 (6.5)
Space group	*I*4
Unit-cell parameters (Å)	$a = b = 96.53$, $c = 56.26$
No. of observed reflections	92613
No. of unique reflections	8542

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \overline{I(hkl)}| / \sum_{hkl}\sum_i I_i(hkl)$ where $I_i(hkl)$ is the *i*th observation of reflection *h* and $\overline{I(hkl)}$ is the mean intensity of all observations of *h*.

This work has been supported by grants from the Ministry of Education of the People's Republic of China (104003) and from the National Natural Science Foundation of China (30325012, 30530190). Peking University's 985 and 211 grants are also gratefully acknowledged. YHL is a recipient of a Fok Ying Tong Education Foundation grant (94017).

References

Barsotti, C. & Ipata, P. L. (2002). *Biochem. Pharmacol.* **63**, 117–122.
Barton, G. J. (1993). *Protein Eng.* **6**, 37–40.
Campobasso, N., Mathews, I. I., Begley, T. P. & Ealick, S. E. (2000). *Biochemistry*, **39**, 7868–7877.
Cheng, G., Bennett, E. M., Begley, T. P. & Ealick, S. E. (2002). *Structure*, **10**, 225–235.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Loesche, W. J. (1986). *Microbiol. Rev.* **50**, 353–380.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Sigrell, J. A., Cameron, A. D., Jones, T. A. & Mowbray, S. L. (1998). *Structure*, **6**, 183–193.
Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). *Nucleic Acids Res.* **25**, 4876–4882.
Zhang, R. G., Grembecka, J., Vinokour, E., Collart, F., Dementieva, I., Minor, W. & Joachimiak, A. (2002). *J. Struct. Biol.* **139**, 161–170.