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

# The structure of the hypothetical protein smu.1377c from *Streptococcus mutans* suggests a role in tRNA modification

**Tian-Min Fu,‡ Xiang Liu,‡ Lanfen Li* and Xiao-Dong Su**

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

‡ These authors contributed equally to this work.

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

Received 18 March 2010
Accepted 20 May 2010

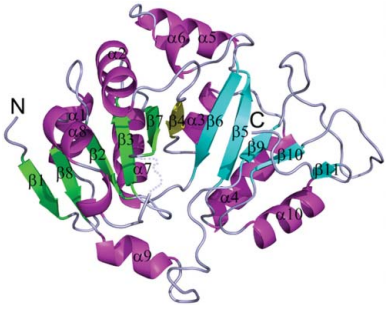
**PDB Reference:** smu.1377c, 3l7v.

Members of the Sua5_YciO_YrdC protein family are found in both eukaryotes and prokaryotes and possess a conserved $\alpha/\beta$ twisted open-sheet fold. The *Escherichia coli* protein YrdC has been shown to be involved in modification of tRNA. The crystal structure of smu.1377c, a hypothetical protein from *Streptococcus mutans*, has been determined to 2.25 Å resolution. From structure analysis and comparison, it is shown that smu.1377c is a member of the Sua5_YciO_YrdC family and that it may play the same role as *E. coli* YrdC.

## 1. Introduction

Members of the Sua5_YciO_YrdC family (Pfam PF01300; Bateman *et al.*, 2002) are widely distributed in eukaryotes and prokaryotes and share a so-called YrdC domain that has been shown to preferentially bind to dsRNA (Teplova *et al.*, 2000; Agari *et al.*, 2008). It has been shown that most organisms possess only one member of the PF01300 family, although some have two or more. YrdC-domain proteins exist either (i) as independent proteins, (ii) with C-terminal extensions or (iii) as domains in larger proteins (http://pfam.sanger.ac.uk/family/PF01300). The *Escherichia coli* YciO and YrdC proteins belong to the first category of independent proteins. The *Saccharomyces cerevisiae* protein Sua5 represents the second category, while *E. coli* HypF is a typical member of the third category (Maier *et al.*, 1996).

Crystal structures of *E. coli* YrdC and YciO are available and reveal an $\alpha/\beta$ twisted open-sheet fold (Teplova *et al.*, 2000; Jia *et al.*, 2002). The structure of *Sulfolobus tokodaii* Sua5 has been reported and the structure of its N-terminal domain is similar to the structures of *E. coli* YrdC and YciO (Agari *et al.*, 2008). The three structures have a positively charged cleft located in the centre that has been suggested to be involved in nucleotide binding (Teplova *et al.*, 2000). Nucleic acid-binding studies of YrdC demonstrated that it preferentially binds double-stranded RNA. Subsequent genetic investigations suggested that YrdC is involved in the maturation of 16S rRNA (Kaczanowska & Ryden-Aulin, 2005). *S. cerevisiae* Sua5 has been identified as a suppressor of a defective translation initiation in the leader region of the iso-1-cytochrome *c* (*cyc1*) gene (Hampsey *et al.*, 1991; Na *et al.*, 1992). These results suggest that the members of the PF01300 family may play a role in translation with pleiotropic effects, although the molecular mechanisms remain unclear. Recently, it was established by a combined approach of bioinformatics, genetics and biochemistry that PF01300 proteins are involved in $N^6$-threonylcarbamoyladenosine modification of the $tRNA^{Lys}$ position A37 ($t^6A$; El Yacoubi *et al.*, 2009). Here, we present the crystal structure of smu.1377c at 2.25 Å resolution. smu.1377c (gi:24379784; GeneID 1028640; 261 amino-acid residues; 29.4 kDa) is one of the selected targets in the *Streptococcus mutans* structural genomics project that is in progress at Peking University (Su *et al.*, 2006). As a hypothetical



protein from *S. mutans*, smu.1377c shows high structural similarity to *E. coli* YrdC, *E. coli* YciO and *S. tokodaii* Sua5 despite having low sequence identity (the sequence identity between smu.1377c and these proteins is 15, 25 and 16%, respectively). Further structural analysis shows that smu.1377c also has a positively charged putative binding cleft and conserved residues at positions that are essential for the activity of the proteins in this family. Based on these structural features, we propose that smu.1377c is a member of the PF01300 family and that this protein may be involved in modification of tRNA in *S. mutans*.

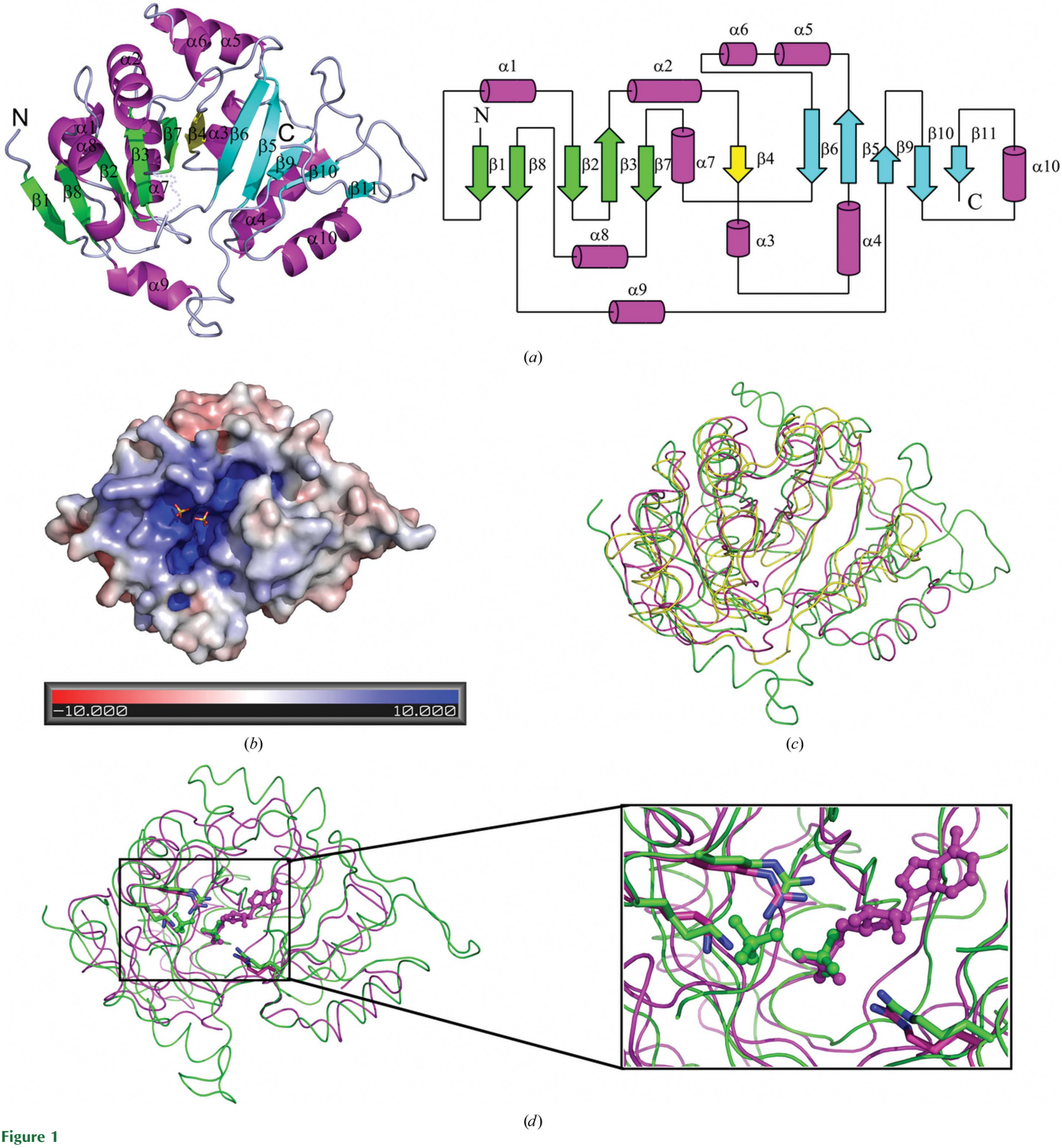



**Figure 1**
(*a*) Ribbon diagram and topology of smu.1377c. $\beta$-Strands of the N-terminal and C-terminal subdomains are coloured green and cyan, respectively. The $\beta$4 strand is shown in yellow. (*b*) Electrostatic potential surface of the smu.1377c structure with the positions of sulfate ions shown as stick models (red and yellow represent O and S atoms, respectively). Red and blue colours represent negative and positive electrostatic potentials ($-10k_BT$ and $+10k_BT$), respectively. (*c*) Structural overlay of smu.1377c (green), *E. coli* YrdC (yellow) and *S. tokodaii* Sua5 (magenta). (*d*) Two sulfate ions (green) bind to smu.1377c (green) and one of them mimics the position of the phosphate group of AMP in the *S. tokodaii* Sua5 structure (magenta). The sulfates and AMP are shown as ball-and-stick models. The side chains of the amino acids involved in AMP and sulfate binding are shown as sticks. This figure was generated in *PyMOL* (DeLano, 2002).

## 2. Materials and methods

### 2.1. Cloning, expression and purification

The *smu.1377c* gene was amplified by polymerase chain reaction (PCR) from genomic *S. mutans* DNA. The PCR product was digested by *Bam*HI and *Xho*I and inserted into the *E. coli* expression vector pET-28a (Novagen). Thus, the gene product was fused with an N-terminal $His_6$ tag with sequence MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGS. The recombinant plasmid was transformed into *E. coli* BL21 (DE3) (Invitrogen) competent cells for expression. Overexpression of smu.1377c was induced with 0.5 m*M* isopropyl $\beta$-D-1-thiogalactopyranoside when the cell density reached an $OD_{600}$ of 0.6–0.8. After further growth for 3 h at 310 K, the cells were harvested by centrifugation and resuspended in 20 ml buffer containing 20 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl. Selenomethionine-substituted protein was produced in *E. coli* BL21 (DE3) using a previously published protocol (Hendrickson *et al.*, 1990).

The resuspended cells were lysed by sonication on ice and cell debris was removed by centrifugation at 3500*g* for 40 min. The supernatant was filtered and loaded onto a 5 ml nickel-chelating column (HiTrap Chelating, GE Healthcare) equilibrated with buffer (20 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl). The column was washed (buffer: 20 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl, 100 m*M* imidazole) and the bound protein was eluted using imidazole (buffer: 20 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl, 300 m*M* imidazole). Fractions containing the target protein were pooled and concentrated to 1 ml by ultrafiltration with a Millipore centrifugal ultrafiltration device (Amicon Ultra, 10 kDa cutoff). Further purification was achieved by gel filtration (Superdex 75, GE Healthcare; buffer: 20 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl).

### 2.2. Protein crystallization and data collection

After gel filtration, smu.1377c fractions were collected and concentrated to about 60 mg ml$^{-1}$ for crystallization. The protein concentration was checked using a Bio-Rad protein-assay kit (Bio-Rad Laboratories, USA) based on the method of Bradford, using bovine serum albumin (BSA) as the standard protein. Initial crystallization screening was carried out at 291 K by the hanging-drop vapour-diffusion method with commercially available crystallization screens from Hampton Research (USA) and XtalQuest (Beijing, China). Cubic shaped crystals appeared within one week in several conditions. After optimization, crystals suitable for diffraction experiments were obtained using 2.0 *M* ammonium sulfate, 0.1 *M* Tris–HCl pH 8.5 with a protein concentration of 40 mg ml$^{-1}$.

For data collection, a single SeMet-derivative crystal (approximate dimensions 0.3 × 0.3 × 0.4 mm) was flash-frozen in mother liquor supplemented with 20% glycerol. A SAD (single-wavelength anomalous diffraction) data set (360 frames, 0.5° oscillation per image) was collected at 100 K on beamline I911-3 at MAX-lab, Sweden. The data were processed with the *XDS* package (Kabsch, 2010) in space group $P2_12_12$, with unit-cell parameters $a$ = 81.9, $b$ = 92.6, $c$ = 40.6 Å. Data-collection statistics are summarized in Table 1.

### 2.3. Structure determination and refinement

The Se sites were located with *SHELXD* (Sheldrick, 2008). The heavy-atom substructure was refined and initial phases were calculated using the *phenix.autosol* program (Adams *et al.*, 2002). The initial model was built using *phenix.autobuild* (Adams *et al.*, 2002) and missing residues were added manually with *Coot* (Emsley & Cowtan, 2004). The final model was refined using *phenix.refine* (Adams *et al.*, 2002) and the stereochemistry of the model was checked by *PROCHECK* (Laskowski *et al.*, 1993) from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). The refinement statistics are summarized in Table 1.

**Table 1**
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

| | |
|---|---|
| Data collection | |
| Wavelength (Å) | 0.97918 |
| Space group | $P2_12_12$ |
| Unit-cell parameters | $a$ = 81.9, $b$ = 92.6, $c$ = 40.6, $\alpha = \beta = \gamma = 90$ |
| Resolution (Å) | 50.00–2.25 (2.30–2.25) |
| $R_{merge}$† (%) | 11.0 (37.3) |
| Mean $I/\sigma(I)$ | 8.85 (2.76) |
| Completeness (%) | 96.9 (71.6) |
| Redundancy | 3.4 (2.0) |
| Refinement | |
| Resolution range (Å) | 50.00–2.25 |
| No. of reflections | 26522 |
| $R_{work}$‡/$R_{free}$§ (%) | 19.9/25.9 |
| Average $B$ factor (Å$^2$) | 29.81 |
| Ramachandran plot (%) | |
| Most favoured | 92.6 |
| Allowed | 7.4 |
| Disallowed | 0.0 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.010 |
| Bond angles (°) | 1.150 |

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of reflection $hkl$ and $\sum_i$ is the sum over all $i$ measurements of reflection $hkl$. ‡ The $R$ factor $R_{work} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where $F_{obs}$ and $F_{calc}$ are the observed and calculated structure factors, respectively. § $R_{free}$ is the $R$ factor calculated over a subset of the data that were excluded from refinement.

## 3. Results and discussion

### 3.1. Overall structure of smu.1377c

We have determined the crystal structure of the *S. mutans* smu.1377c protein at 2.25 Å resolution by SAD phasing; each asymmetric unit contained one molecule of smu.1377c. The structure was refined to reasonable $R$ values ($R_{work}$ = 19.9%; $R_{free}$ = 25.9%). Residues 157–161 were not included in the final model owing to poor electron density.

The structure of smu.1377c consists of ten $\alpha$-helices and 11 $\beta$-strands and folds as an $\alpha/\beta$ twisted open sheet (Fig. 1*a*). The overall structure of smu.1377c can be divided into two subdomains, each containing five $\beta$-strands and five $\alpha$-helices: $\beta1$–$\beta2$–$\beta3$–$\beta7$–$\beta8$ and $\alpha1$–$\alpha2$–$\alpha7$–$\alpha8$–$\alpha9$ constitute the N-terminal subdomain, while the other strands and helices belong to the C-terminal subdomain.

The electrostatic potential surface of smu.1377c revealed that a positively charged concave cleft is located in the centre of the structure (Fig. 1*b*). Two sulfate ions, stabilized by water molecules and positive residues, were observed to be bound to this cleft in the smu.1377c structure (Fig. 1*b*).

### 3.2. Structure comparison

No homologues with high sequence identity to smu.1377c in the Protein Data Bank (PDB) were found by searching with *BLAST* (Altschul *et al.*, 1990), but a search for structural homologues using *DALI* (Holm & Sander, 1993) revealed four unique hits with a $Z$ score (similarity score) of above 10. The properties of these four structural homologues are shown in Table 2. Both *E. coli* YrdC and *S. tokodaii* Sua5 are among the hits. The structure of smu.1377c can be superimposed on those of *E. coli* YrdC and the N-terminal domain of *S. tokodaii* Sua5 with r.m.s.d.s of 2.7 and 2.1 Å, respectively (Fig. 1*c*

**Table 2**
The properties of four structural homologues of smu.1377c.

All four structural homologues were found from the PDB using the program *DALI*.

| Protein | Source | PDB code | *Z* score | R.m.s.d. value (Å) | Sequence identity to smu.1377c (%) | Overlapped amino-acid residues | Reference |
|---|---|---|---|---|---|---|---|
| Sua5 | *S. tokodaii* | 2eqa | 20.4 | 2.1 | 16 | 34/209 | Agari *et al.* (2008) |
| YciO | *E. coli* | 1kk9 | 18.1 | 2.4 | 25 | 46/186 | Jia *et al.* (2002) |
| YrdC | *E. coli* | 1hru | 14.6 | 2.7 | 15 | 29/190 | Teplova *et al.* (2000) |
| MTH1692 | *Methanobacterium thermoautotrophicum* | 1jcu | 10.2 | 3.1 | 24 | 49/208 | Yee *et al.* (2002) |

and Table 2). The overlaid structures of smu.1377c, *E. coli* YrdC and the N-terminal domain of *S. tokodaii* Sua5 depict that despite low sequence identity to each other these structures share a similar fold (Fig. 1*c*). All of the structures contain a concave positively charged cleft, which has been shown to provide nucleic acid-binding capacity to YrdC (Teplova *et al.*, 2000), and we propose that this area of smu.1377c may also be involved in nucleic acid binding. Previous investigations revealed the involvement of YrdC in $t^6A$ modification and identified two residues, Lys50 and Arg52 in YrdC, as being essential for this specific YrdC function (El Yacoubi *et al.*, 2009). Sequence alignment showed that these two residues were strictly conserved among the members of the family (Lys51 and Arg53 in smu.1377c; Fig. 2). This feature suggests that smu.1377c may also be involved in $t^6A$ tRNA modification. In addition, one of the two adjacent sulfate ions in the conserved concave cleft of the smu.1377c structure superimposes well on the phosphate group of the AMP located in the *Su. tokadaii* Sua5 structure (Fig. 1*d*). Thus, the two sulfate ions in our structure may mimic the phosphate groups of ATP, which has been shown to be a required cofactor for *E. coli* YrdC and *Su. tokadaii* Sua5. Genomic analysis shows that smu.1377c is the one and only member of the PF01300 family in *S. mutans*. These data further indicate the role of smu.1377c in tRNA modification.

In addition to the structural similarity, smu.1377c also shows some unique features that distinguish it from *E. coli* YrdC and the N-terminal domain of *S. tokodaii* Sua5. smu.1377c is much longer (by 60–80 amino acids). In addition to some small loop insertions, smu.1377c possesses three long insertions and a 20-amino-acid C-terminal extension compared with *E. coli* YrdC and the N-terminal domain of *S. tokodaii* Sua5 (Fig. 2). The functional roles of these features require further investigation.

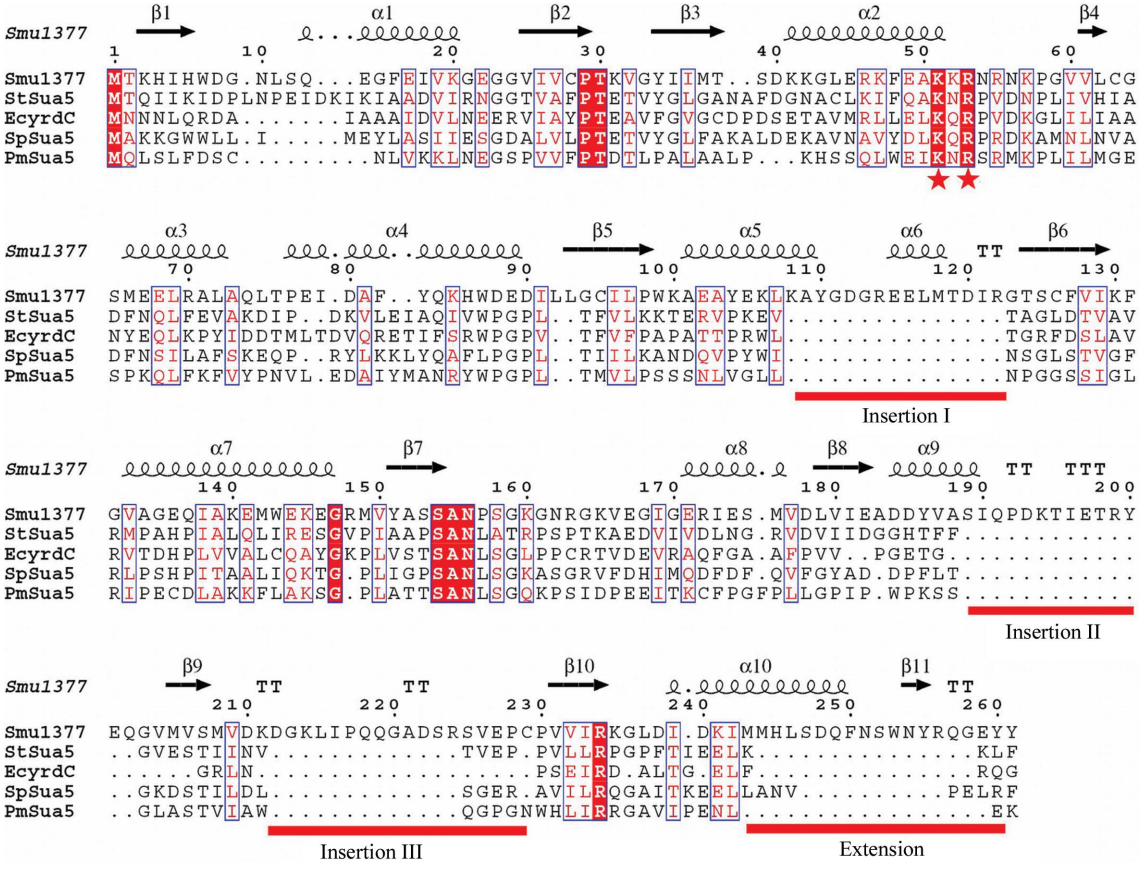



**Figure 2**
Sequence alignment of smu.1377c with *S. tokodaii* Sua5 (StSua5), *E. coli* YrdC (EcYrdC), *S. pyogenes* Sua5 (SpSua5) and *Prochlorococcus marinus* Sua5 (PmSua5). Strictly conserved residues are boxed in red and similar residues are represented by red letters. Residues essential for the function of the YrdC proteins are marked with red asterisks. The insertions and extension are marked by red lines. The alignment was performed using the programs *T-Coffee* (Notredame *et al.*, 2000) and *ESPript* v.2.2 (Gouet *et al.*, 1999). All protein sequences are from the NCBI database, with the following accession codes: *S. mutans* smu.1377c, gi:24379784; *S. tokodaii* Sua5, gi:15921819; *E. coli* YrdC, gi:2851671; *S. pyogenes* Sua5, gi:50903281; *P. marinus* Sua5, gi:159902895.

In summary, the crystal structure of smu.1377c shows that it is a member of the PF01300 family and suggests that it may play a role in the modification of tRNA. Our structure has provided useful clues for further functional study of smu.1377c.

We are grateful to Dr Erik Brostromer for assistance in data collection. We are also grateful to Dr Ole Kristensen for reading and comments on the manuscript. This work was supported by grant No. 30530190 from the National Natural Science Foundation of China.

## References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.
Agari, Y., Sato, S., Wakamatsu, T., Bessho, Y., Ebihara, A., Yokoyama, S., Kuramitsu, S. & Shinkai, A. (2008). *Proteins*, **70**, 1108–1111.
Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. & Sonnhammer, E. L. (2002). *Nucleic Acids Res.* **30**, 276–280.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
DeLano, W. L. (2002). *The PyMOL Molecular Viewer.* http://www.pymol.org.
El Yacoubi, B., Lyons, B., Cruz, Y., Reddy, R., Nordin, B., Agnelli, F., Williamson, J. R., Schimmel, P., Swairjo, M. A. & de Crecy-Lagard, V. (2009). *Nucleic Acids Res.* **37**, 2894–2909.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Hampsey, M., Na, J. G., Pinto, I., Ware, D. E. & Berroteran, R. W. (1991). *Biochimie*, **73**, 1445–1455.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Jia, J., Lunin, V. V., Sauvé, V., Huang, L.-W., Matte, A. & Cygler, M. (2002). *Proteins*, **49**, 139–141.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kaczanowska, M. & Ryden-Aulin, M. (2005). *Biochim. Biophys. Acta*, **1727**, 87–96.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Maier, T., Binder, U. & Bock, A. (1996). *Arch. Microbiol.* **165**, 333–341.
Na, J. G., Pinto, I. & Hampsey, M. (1992). *Genetics*, **131**, 791–801.
Notredame, C., Higgins, D. G. & Heringa, J. (2000). *J. Mol. Biol.* **302**, 205–217.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Su, X.-D., Liang, Y., Li, L., Nan, J., Bröstromer, E., Liu, P., Dong, Y. & Xian, D. (2006). *Acta Cryst.* D**62**, 843–851.
Teplova, M., Tereshko, V., Sanishvili, R., Joachimiak, A., Bushueva, T., Anderson, W. F. & Egli, M. (2000). *Protein Sci.* **9**, 2557–2566.
Yee, A. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 1825–1830.