

Protein Preparation, Crystallization and Preliminary X-ray Crystallographic Analysis of Smu_195c From Caries Pathogen *Streptococcus mutans**

GAO Zeng-Qiang¹⁾, HOU Hai-Feng¹⁾, LI Li-Qin¹⁾, XU Rui¹⁾,
LIANG Yu-He^{2,3)}, LI Lan-Fen^{2,3)}, SU Xiao-Dong^{2,3)}, DONG Yu-Hui^{1)**}

(¹⁾Beijing Synchrotron Radiation Facility, Institute of High Energy Physics, Beijing 100049, China;

²⁾National Laboratory of Protein Engineering and Plant Genetic Engineering, Peking University, Beijing 100871, China;

³⁾Department of Biochemistry and Molecular Biology, College of Life Science, Peking University, Beijing 100871, China)

Abstract Smu_195c is a protein with 86 amino acids in *Streptococcus mutans*, a primary pathogen for human dental caries. The specific function of Smu_195c is still unknown and there are no conserved domains in it. In order to find out its function, the gene encodes Smu_195c was cloned and expressed in *E. coli* as N-terminally 6*His tagged recombinant protein. Two crystal forms were obtained by the hanging drop method. Form I belongs to space group P6₁22 or P6₅22 with the unit cell parameters $a = b = 62.93 \text{ \AA}$, $c = 90.63 \text{ \AA}$, $\gamma = 120^\circ$ and form II belongs to the space group P4₁2₁2 or P4₃2₁2 with the unit cell parameters $a = b = 57.97 \text{ \AA}$, $c = 103.51 \text{ \AA}$. Crystals from the protein with His-tag belong to form I, however, crystals from the protein without His-tag belong to both.

Key words *Streptococcus mutans*, Smu_195c, crystallization, X-ray diffraction

Dental caries is one of the most common infectious disease afflicting humans, and tends to remain untreated in many undeveloped areas because of high cost, leading to considerable suffering that is often alleviated only by the loss or extraction of the inflected tooth^[1,2]. Although 200 to 300 bacterial species have been found associated with dental plaque, only the presence of *S. mutans* has been consistently linked with the formation of human dental caries^[3]. The highest possibility of *S. mutans* serotype c strain being found in humans is 80%^[4]. For all kinds of reasons, the research of the nosogenesis of *S. mutans* is still staying on the degree of gene sequence and structure analysis. There are hardly reports about the protein 3D structure. We and our collaborator chose ATCC25175 (American Type Culture Collection) for the research of structural genomic of the proteins from the whole genome. We clone and express the important genes for dental caries and related proteins cosmically and then analyse its 3D structures. It will be helpful for the further research of the emergence, development and defense of the dental caries.

Smu_195c is one of the proteins expressed from *S. mutans* serotype c strain, although its specific function is unknown, the primary sequence alignment indicates that it is related with some hypothetical phage proteins (Figure 1). We have obtained the crystals and the structure determination is underway.

1 Gene cloning and expressing

To construct expression plasmid, target DNA fragment was amplified by polymerase chain reaction (PCR) from the *S. mutans* genomic DNA, using primers Smu195c-F (5' cgcgatccatgttaacaaaagaaaa-tgtgc 3') and Smu195c-R (5' ccgctcgagtcaaacttgctc-atacgtcgc 3'), which contains *Bam*H I and *Xho* I digested fragments, respectively. The PCR-amplified

*This work was supported by grants from The Nation Foundation of Talent Youth (30325012), The National Natural Science Foundation of China (30530190, 10570113).

**Corresponding author. Tel: 86-10-88233090, Fax: 86-10-88233201, E-mail: dongyh@ihep.ac.cn

Received: July 17, 2006 Accepted: August 26, 2006

fragment was then ligated to the expression vector pET28a (Novagen) by conventional cloning method. The final vector was verified by DNA sequencing. The

constructed plasmid was transformed into *Escherichia coli* strain BL21 (DE3) cells for expression experiments^[5].

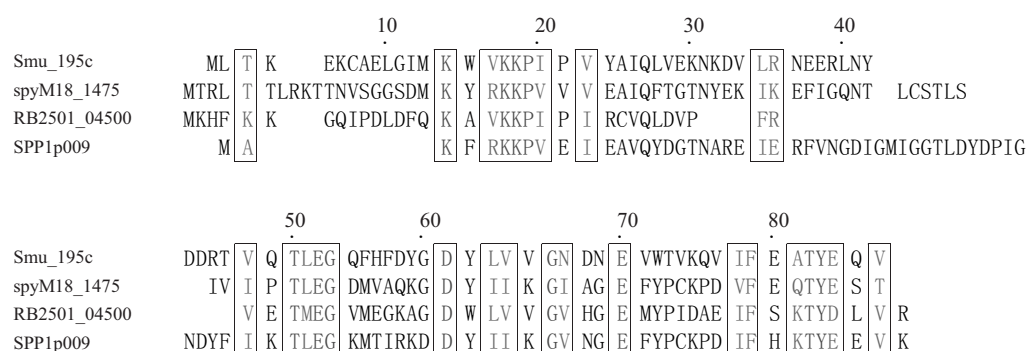


Fig. 1 Primary sequences alignment among smu_195c and other proteins

spyM18_1475: hypothetical phage protein from *Streptococcus pyogenes* MGAS8232; RB2501_04500: hypothetical protein from *Robiginitalea biformata* HTCC2501; SPP1p009: Hypothetical 10.3 ku protein in GP2-GP6 intergenic region from *Bacillus* phage SPP1. The sequence alignment was performed using the program CLUSTALX^[6] and ESPRIPT^[7].

2 Protein expression and purification

The protein was expressed in *E.coli* strain BL21 (DE3) as follows. A single clone including the Smu_195c gene was inoculated to 50 ml LB with 50 mg/L kanamycin and cultured at 310 K overnight (about 10 h). The overnight culture was transferred into 950 ml LB with 50 mg/L kanamycin and continued to grow at 310K until the A_{600} value reached 0.6~0.8. Isopropyl-1-thio- β -D-galactopyranoside (IPTG) at a final concentration of 0.4 mmol/L was added and cultured for another 4 h. The cells were collected by centrifugation at 6 000 r/min for 10 min at 277 K. The pellet was resuspended and sonicated in buffer A (20 mmol/L Tris-HCl, 500 mmol/L NaCl pH 7.5). The lysate was centrifugated at 1 6000 r/min for 45 min at 277 K, the supernatant was filtered by 0.22 μ m filter and then applied to a 5 ml Hi-Trap Ni column (Amersham) equilibrated with buffer A. The column was washed with buffer B (500 mmol/L imidazole contained in buffer A). The fraction was checked by SDS-PAGE (Figure 2) and those containing the target protein were pooled and concentrated in an Amicon ultrafiltration concentrator with a 10 ku molecular-mass cutoff. The concentrated protein was applied to a Hi-Load Superdex-75 column (Amersham) equilibrated with buffer C (20 mmol/L Tris-HCl, 100 mmol/L NaCl, pH 7.5). The peak fraction was pooled and concentrated in the same concentrator as before. The concentration of the protein was measured by a

Bio-Rad Protein Assay kit and the final concentration is about 60 g/L

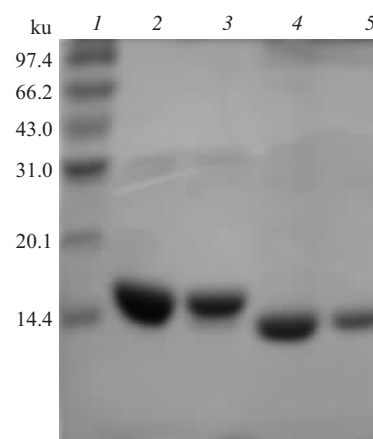


Fig. 2 SDS-PAGE of Smu_195c

1: Molecular mass marker; 2: The protein purified by Ni column with his-tag; 3: The protein purified by superdex-75 with his-tag; 4: The same as 3, but without his-tag; The protein from column 4 was purified by superdex-75 again, and the result shows in column 5.

3 Crystallization and data collection

The protein was diluted to 30 g/L with buffer C for crystallization. Hanging-drop method was used. The initial screening was performed using Hampton Research Crystal Screen kit I, II and Index (Hampton Research, Riverside, CA, USA) at 293 K. Each drop contains 1 μ l protein and 1 μ l reservoir solution and equilibrated against 500 μ l reservoir solution in the well. Needle-like and diamond-like

crystals appeared from many conditions (the former mainly from Index kit and the later from Screen I kit). After optimization, crystals suitable for X-ray diffraction were obtained under two kinds of conditions. One including 25% PEG1000, 0.1 mol/L HEPES pH 7.7 and 0.2 mol/L MgCl_2 or CaCl_2 , another including 25% PEG1000, 0.2 mol/L HEPES pH 7.7 and 0.05 mol/L MgCl_2 . The later was interesting, crystals from the protein with His-tag belong to the space group $P6_122$ or $P6_522$ with the unit-cell parameters $a = b = 62.93 \text{ \AA}$, $c = 90.63 \text{ \AA}$, $\gamma = 120^\circ$, but crystals from the protein without His-tag (the his-tag was cut off by thrombin as follows: after being purified by Superdex-75 column, the protein was mixed with thrombin in the ratio 60 μg per unit in 3 ml buffer C, staying for about 20 h at the room temperature, and then purified by Superdex-75 column again. The thrombin was purchased from Ding-guo company) belong to the space group $P4_12_12$ or $P4_32_12$ with the unit-cell parameters $a = b = 57.97 \text{ \AA}$, $c = 103.51 \text{ \AA}$ (Figure 3).

X-ray diffraction data of crystals with and without His-tag were collected on 3W1A of BSRF (Beijing Synchrotron Radiation Facility) and Rigaku 007, respectively. Data were processed and scaled with the software HKL2000^[8], the results were listed in Table 1.

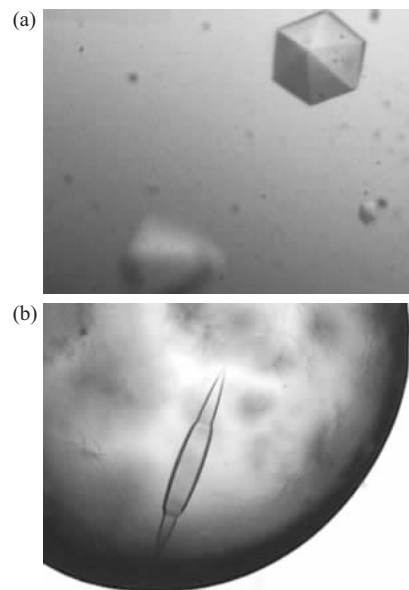


Fig. 3 Crystals of Smu_195c
(a)form I, (b)form II.

We also obtained the Se-Met crystals which could diffract to $2.6 \text{ \AA}$, and we were sure that there were some Se atoms in the crystal by fluorescence signal, however, we failed to locate the Se atoms (by the software SOLVE^[9], SHELXD^[10] and PHENIX^[11]). One possibility is that the Se atoms lie in the loop region. The structure determination is still underway.

Table 1 X-ray data collection statistics

	Form I	Form II
X-ray source	BSRF	Rigaku 007
Wavelength / $\text{\AA}$	1.0	1.5418
Space group	$P6_122$ or $P6_522$	$P4_12_12$ or $P4_32_12$
Unit cell parameter / $\text{\AA}$	$a = b = 62.93$, $c = 90.63$	$a = b = 57.97$, $c = 103.51$
Resolution / $\text{\AA}$	$50 \sim 2.4$ ($2.49 \sim 2.40$) ¹⁾	$18 \sim 2.6$ ($2.7 \sim 2.6$)
Rsym / % ²⁾	6.9 (33.2)	6.6 (36.3)
I/σ	25 (4)	30 (6)
Unique reflections	4 503	6 527
Observed reflections	3 2016	58 175
Completeness / %	99.0	99.8
Mol/Asym	1	1 or 2
$V_m(\text{\AA}^3 \text{ Da}^{-1})$ ^[12]	2.0	4.0 or 2.0
Solvent content / %	36.8	68.6 or 37.3

¹⁾ Data in parentheses corresponds to the highest resolution shell. ²⁾ $\text{Rsym} = \sum |I_{\text{obs}} - I_{\text{avg}}| / \sum I_{\text{obs}}$ where the summation is over all reflections.

Acknowledgment We would like to thank Professor Wei-Min Gong of Institute of Biophysics of The Chinese Academy of Sciences for valuable advice of structure determination.

References

- 1 W.H.O., Geneva. Fluorides and Oral Health. In: World Health Organization Report. 1994
- 2 Ajdic D, McShan W M, McLaughlin R E, *et al.* Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. Proc Natl Acad Sci USA, 2002, **99** (22): 14434~14439
- 3 Loesche W J. Role of *Streptococcus mutans* in human dental decay. Microbiol Rev, 1986, **50**: 353~380
- 4 刘天佳. 口腔疾病的微生物学基础. 北京: 人民卫生出版社, 1999
Liu T J. The Microbiology Basis of Nonnasality Illness. Beijing: People Health Press, 1999. 17~19
- 5 Duan M R, Ren H, Mao P, *et al.* Crystallization and preliminary X-ray analysis of the C-terminal WRKY domain of *Arabidopsis thaliana* WRKY1 transcription factor. Biochim Biophys Acta, 2005, **1750**: 14~16
- 6 Thompson J D, Gibson T J, Plewniak F, *et al.* The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 1997, **24**: 4876~4882
- 7 Gouet P, Courcelle E, Stuart D I, *et al.* ESPript: multiple sequence alignments in PostScript. Bioinformatics, 1999, **15**: 305~308
- 8 Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. methods in enzymology, 1997, **276**: 307~326
- 9 Terwilliger T C, Berendzen J. Automated MAD and MIR structure solution. Acta Cryst D, 1999, **55**: 849~861
- 10 Schneider T R, Sheldrick G M. Substructure solution with *SHELXD*. Acta Cryst D, 2002, **58**: 1772~1779
- 11 Adams P D, Grosse-Kunstleve R W, Hung L W, *et al.* PHENIX: building new software for automated crystallographic structure determination. Acta Cryst D, 2002, **58**: 1948~1954
- 12 Matthews B W. Solvent content of protein crystals. J Mol Biol, 1968, **33** (2): 491~497

龋齿致病菌变性链球菌蛋白 Smu_195c 的 制备、结晶及初级晶体学分析*

高增强¹⁾ 侯海峰¹⁾ 李丽琴¹⁾ 许锐¹⁾
梁宇和^{2,3)} 李兰芬^{2,3)} 苏晓东^{2,3)} 董宇辉^{2,3)}

¹⁾中国科学院高能物理研究所同步辐射实验室, 北京 100049;

²⁾北京大学生命科学学院, 蛋白质工程和植物基因工程国家实验室, 北京 100871;

³⁾北京大学生命科学学院生物化学与分子生物学系, 北京 100871)

摘要 变性链球菌(*Streptococcus mutans*)是最主要的龋齿致病菌. 其中 Smu_195c 为一个约 10 ku (86 个氨基酸) 的蛋白质, 序列比对结果显示, 它没有包含保守的结构域, 功能也是未知的. 为揭示其功能, 在大肠杆菌 BL21 (DE3) 中表达了 N 端结合了 6 个 His 的编码 Smu_195c 的 DNA 片段. 通过悬滴法得到了两种不同晶系的晶体. 第一种属于空间群 P6₁22 或 P6₂22, 晶格参数为 $a=b=62.93\text{ \AA}$, $c=90.63\text{ \AA}$, $\gamma=120^\circ$; 第二种属于空间群 P4₁2₁2 或 P4₃2₁2, 晶格参数为 $a=b=57.97\text{ \AA}$, $c=103.51\text{ \AA}$. N 端带有 His 的蛋白质长出的晶体属于第一种, 而切掉 N 端的 6 个 His 后的蛋白质长出的晶体两种都有.

关键词 变性链球菌, Smu_195c, 结晶, X 射线衍射

学科分类号 Q51, O72

*国家自然科学基金杰出青年基金(30325012), 国家自然科学基金重点项目“致龋变形链球菌全基因组的蛋白表达及三维结构研究”(30530190)和“利用同步辐射综合方法研究变形链球菌未知蛋白的结构和功能关系”(10575113)资助.

** 通讯联系人. Tel: 010-88233090, Fax: 010-88233201, E-mail: dongyh@ihep.ac.cn

收稿日期: 2006-07-17, 接受日期: 2006-08-26