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

Overexpression, purification, crystallization and data collection of *Sulfolobus solfataricus* Sso6206, a novel highly conserved protein

Andrew R. McEwan, Huanting Liu, Muse Oke, Lester Carter, Helen Powers, Mark Dorward, Stephen A. McMahon, Malcolm F. White and James H. Naismith*

The Centre for Biomolecular Sciences, The University, St Andrews KY16 9ST, Scotland

Correspondence e-mail: naismith@st-and.ac.uk

Received 16 November 2005
Accepted 30 January 2006

Sso6206, a 10.5 kDa protein from *Sulfolobus solfataricus*, has been overexpressed, purified and crystallized. The protein crystallizes in space group $P6_{1/5}22$, with unit-cell parameters $a = b = 157.8$, $c = 307.3$ Å. The crystals are hexagonal bipyramids and a data set has been collected to 2.4 Å resolution. Molecular replacement cannot be attempted as no convincing model can be identified. Crystals of selenomethionine-variant protein have not yet been obtained. Interestingly, crystal packing, gel filtration and mass spectrometry all suggest the native protein forms a multi-subunit oligomer consisting of >9 subunits.

1. Introduction

Sso6206 is a small acidic protein (MW 10 483 Da, pI 4.5) encoded in the genome of the crenarchaeon *Sulfolobus solfataricus* (She *et al.*, 2001) with an unknown function. It is nearly ubiquitous in the archaea, including crenarchaea and euryarchaea, hyperthermophiles, halophiles, psychrophiles and methanogens. In addition, homologues have been annotated in two bacterial species, *Shewanella frigidimarina*, a marine member of the gamma proteobacteria (Bowman *et al.*, 1997), and *Desulfotalea psychrophila*, a sulfate-reducing delta proteobacterium from arctic sediments (Knoblauch *et al.*, 1999). The limited distribution in bacteria suggests that these two species have obtained the gene by lateral gene transfer from an archaeon. The amino-acid sequence of the protein is highly conserved (Fig. 1), with greater than 40% identity observed between archaeal family members. Perhaps significantly, there are four conserved acidic residues and a conserved histidine (coloured red and blue, respectively, in Fig. 1), which are often ligands for bound metal ions. The strong conservation hints at a role for this protein in the archaea, making it an interesting target for structural studies. The protein belongs to Pfam2680, a family which has yet to be characterized. Here, we report the overexpression, crystallization and native data collection of recombinant Sso6206.

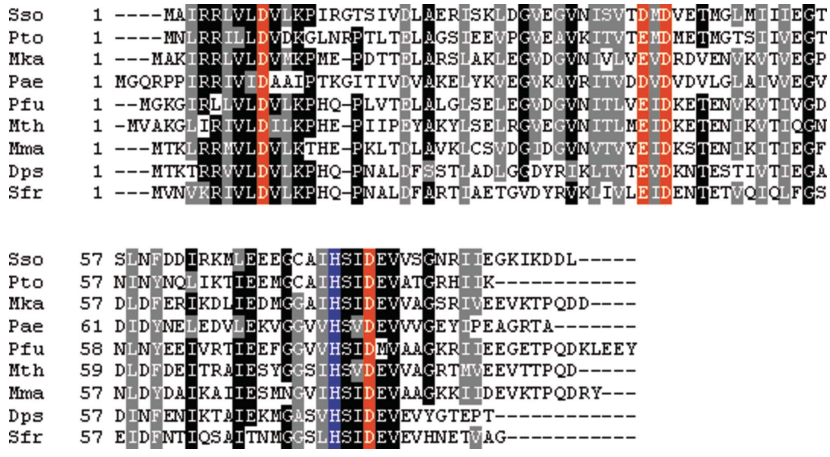
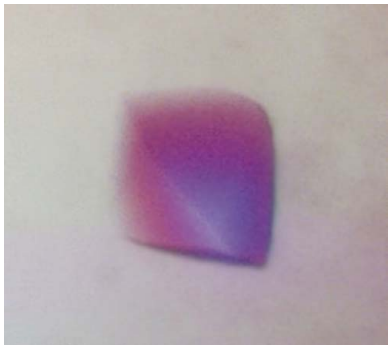


Figure 1
Sequence alignment of Sso6206 and its homologues from archaea and bacteria. The four conserved acidic residues are highlighted in red and the conserved histidine in blue. Sso, Sso6206 from *S. solfataricus*; Pto, Mka, Pae, Pfu, Mth, Mma, Dps and Sfr, homologues from *Picrophilus torridus*, *Methanopyrus kandleri*, *Pyrobaculum aerophilum*, *Pyrococcus furiosus*, *Methanothermobacter thermautotrophicum*, *Methanococcus maripaludis*, *Desulfotalea psychrophila* and *Shewanella frigidimarina*, respectively. Sso6206 contains a single cysteine residue.



2. Materials and methods

2.1. Cloning, overexpression and purification

The Sso6206 gene was amplified using the genomic DNA of *Sulfolobus solfataricus* P2 as a template with a 5′-end oligonucleotide primer (5′-GGCGCCATGGATGGCAATTAGAAGACTTGTTT-TAG-3′) containing an *Nco*I site and a 3′-end primer (5′-GCTG-GGATCCCTACAAATCATCCTTTATCTTTCCCTC-3′) containing a *Bam*HI site. The amplified DNA fragment was digested with *Nco*I/*Bam*HI restriction enzymes and then ligated into an *Nco*I/*Bam*HI linearized pEHISTEV vector (Liu & Naismith, unpublished work) such that six histidine residues and a tobacco etch virus (TEV) protease site were added to the N-terminus of the protein in order to facilitate automated protein purification. After protease cleavage, the resulting protein has an additional glycine and alanine at the N-terminus.

The recombinant plasmid was transformed into *Escherichia coli* BL21 (DE3) cells (Novagen) and four 500 ml cultures were grown in Luria–Bertani medium (Formedia) supplemented with 30 µg ml^{-1} kanamycin at 310 K until A_{600} reached 0.6. The cultured cells were induced with 1 m*M* isopropyl β-D-thiogalactopyranoside (IPTG) and incubated for 4 h at 298 K. The cells were harvested by centrifugation (5000 rev min^{-1}, 15 min, 277 K), washed with PBS and re-spun. The cells were stored at 193 K until required. For purification, the cell pellet was resuspended in lysis buffer (25 m*M* Tris–HCl pH 7.75, 0.6 *M* NaCl, 30 m*M* imidazole, 20 µ*M* lysozyme and 20 µg ml^{-1} DNAse I) and lysed by a Constant Systems cell disruptor at 298 K. The crude lysate was centrifuged (15 000 rev min^{-1}, 25 min, 277 K) and the supernatant fraction was passed through a 0.22 µm PES syringe filter (Millipore). The supernatant was passed over a nickel-affinity column and the protein eluted in a single step with elution buffer (25 m*M* Tris–HCl pH 7.75, 0.6 *M* NaCl, 250 m*M* imidazole). After removal of the imidazole, the eluted protein was then incubated with 1 ml of 1 mg ml^{-1} tobacco etch virus protease (approximate 20:1 ratio protein:protease). The protein was then applied onto an S-200 size-exclusion chromatography column and fractions were assayed by SDS–PAGE gel electrophoresis. The His-tagged and cleaved protein behave similarly during gel filtration and the presence or absence of 20 m*M* DTT does not significantly alter the elution profile of either protein. The fractions containing homogeneous Sso6206 were combined. It was estimated that a total of 20 mg pure Sso6206 was obtained from 2 l cell culture. The protein was concentrated to 6 mg ml^{-1} as measured by Bradford assay. The identity and integrity of the protein were confirmed by mass spectrometry (University of St Andrews).

2.2. Crystallization

Initial conditions were obtained using a sitting-drop vapour-diffusion screen of commercial sparse-matrix crystallization conditions. Three protein concentrations (6, 3 and 1 mg ml^{-1}) were screened at 293 K with a drop size of 0.2 µl (containing 0.1 µl protein solution and 0.1 µl precipitant solution) prepared using a nanodrop crystallization robot (Cartesian HoneyBee) as a part of the Hamilton–Thermo Rhombix system. Optimization of the initial 12 most promising hits, including at least one at each concentration, were performed to confirm that the crystals contained protein. The largest protein crystals were obtained using the hanging-drop vapour-diffusion method (2 µl + 2 µl) at 293 K. Improved crystals could be obtained from a number of conditions; however, the best crystals, as judged by size and regular shape, were obtained from variation of Wizard I (Emerald Biosystems) condition Nos. 31 [20%(*w*/*v*) PEG 8000, 0.1 *M* phosphate–citrate pH 4.2, 0.2 *M* NaCl] and 36 (1.0 *M* sodium citrate, 0.1 *M* imidazole pH 8.0). The crystals obtained from both conditions were of the protein and had hexagonal bipyramidal morphology (0.4 × 0.4 × 0.4 mm). The crystals appear as a spherical precipitate within 2 d and gradually develop more distinct edges over the course of a month. Diffraction analysis showed that the crystals obtained from a precipitant of 0.6 *M* sodium citrate, 0.1 m*M* imidazole pH 7.75 with a protein concentration of 5 mg ml^{-1} were optimal (Fig. 2). We have been able to express the selenomethionine variant of the protein using methioinine inhibition (Doublié, 1997). The protein was purified in the same way as the native protein. Mass spectrometry confirmed full incorporation of selenium and did not indicate that any oxidation had occurred. The selenomethionine protein appeared to be homogenous by mass spectrometry and SDS–PAGE gel electrophoresis. However, we have so far been unable to crystallize the selenomethionine variant of the protein either using

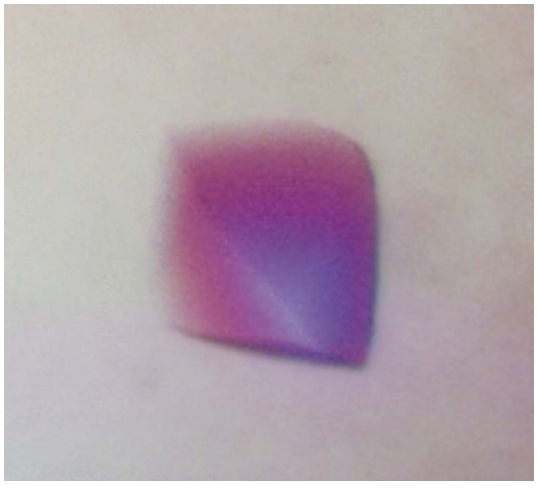


Figure 2
Crystal of Sso6206.

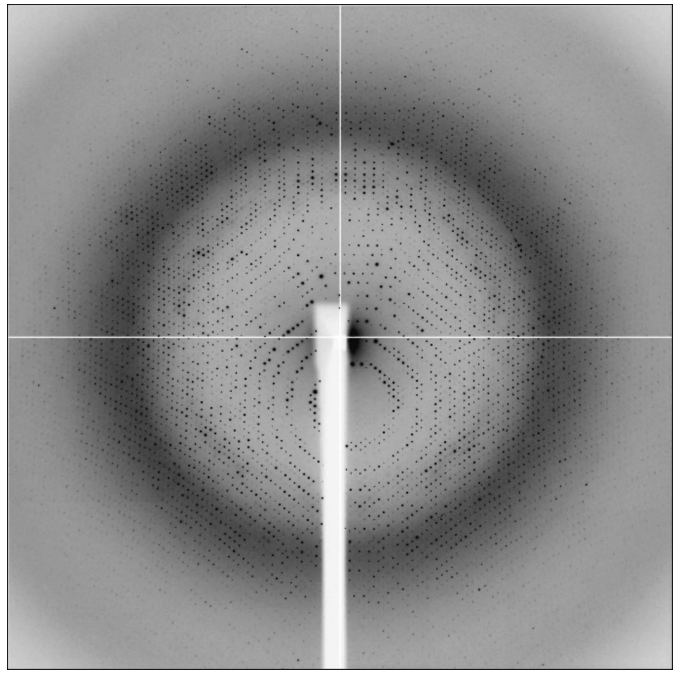


Figure 3
Diffraction pattern of Sso6206 collected at the ESRF.

Table 1
Crystal data and data-collection statistics.

Values in parentheses refer to the highest resolution shell.

	ESRF	SRS	In-house
Wavelength (Å)	0.98	0.87	1.54
Resolution (Å)	76.47–2.40 (2.53–2.40)	69.0–2.50 (2.60–2.50)	45.13–2.80 (2.95–2.80)
Space group	$P6_{1/5}22$	$P6_{1/5}22$	$P6_{1/5}22$
Temperature (K)	130	130	130
Detector	MAR 225 CCD	ASDC Q4 CCD	Rigaku HTC IP
Unit-cell parameters (Å, °)	$a = b = 157.8$, $c = 307.3$, $\alpha = \beta = 90$, $\gamma = 120$	$a = b = 155.2$, $c = 302.7$, $\alpha = \beta = 90$, $\gamma = 120$	$a = b = 158.1$, $c = 308.3$, $\alpha = \beta = 90$, $\gamma = 120$
Solvent content (%)	49.10	46.25	49.12
Unique reflections	89232 (12793)	59867 (8561)	56502 (8126)
$I/\sigma(I)$	26.6 (8.0)	28.5 (6.6)	19.7 (6.6)
Average redundancy	15.8 (16.1)	12.6 (12.9)	8.4 (8.5)
Data completeness (%)	100.0 (100.0)	100.0 (100.0)	99.8 (100.0)
$R_{\rm merge}$†	0.085 (0.356)	0.084 (0.367)	0.098 (0.303)

† $R_{\rm merge} = \sum\sum I(h)_i - \langle I(h)\rangle / \sum I(h)_i$, where $I(h)$ is the measured diffraction intensity and the summation includes all observations.

the native conditions or in a re-screen of sparse-matrix conditions. Circular-dichroism spectrometry suggests the selenomethionine protein is not significantly different and the protein has a similar profile to the native protein on size-exclusion chromatography.

2.3. X-ray data collection

A 2.7 Å data set was collected in-house at 130 K from a single crystal mounted on a loop. The crystal was subject to a soaking in 15%(*v*/*v*) ethylene glycol in stabilization buffer before transferring to cryogenic conditions for data collection. Data was collected as 20 min 0.5° oscillations using a Rigaku HTC image-plate detector, a rotating-anode X-ray source and Osmic mirrors. The crystal-to-detector distance was 220 mm. All data were integrated using *MOSFLM* (Leslie, 1992) and merged with *SCALA* (Evans, 1997) as implemented in *CCP*4 (Collaborative Computational Project, Number 4, 1994). The initial indexing showed the crystals to be primitive hexagonal, with unit-cell parameters $a = b = 158.1$, $c = 308.3$ Å, $\alpha = \beta = 90$, $\gamma = 120°$; a second native data set was collected at Daresbury which showed similar space group and resolution. A third data set was collected at the European Scientific Research Facility (ESRF) Grenoble with the same space group and unit-cell parameters and improved resolution (Fig. 3). Data-collection statistics are given in Table 1. Merging of the data indicates that the space group belongs to the higher symmetry 6/*mmm* Laue class. Analysis of over 100 00*l* axial reflections shows a clear $l = 6n$ condition, indicating that the space group is $P6_122$ or its enantiomorph $P6_522$.

3. Results and discussion

There is no convincing sequence match between Sso6206 and any structure in the PDB. This rules out molecular replacement as a phasing strategy. Several attempts to incorporate heavy atoms by soaking into existing crystals or by cocrystallization with recombinant protein have proven unsuccessful in that the crystals which grew or survived soaking did not show a clear anomalous diffraction when tested at the SRS Daresbury (λ = 1.488 Å, data not shown). We will continue with heavy-metal soaks and attempt to use sulfur phasing to obtain a structure.

The unit-cell parameters are consistent with an asymmetric unit content of between 15 ($V_{\rm M}$ = 3.50 Å^3 Da^{-1}, solvent content 65%) and 30 ($V_{\rm M}$ = 1.77 Å^3 Da^{-1}, solvent content 30%) monomers. The relatively low diffraction resolution of the crystals suggests a higher solvent content. Interestingly, the protein elutes very far from the expected elution point for a 10 kDa protein on a Sephacryl S-200 column. On this column globular proteins in excess of 200 kDa elute in the void volume and Sso6206 does not elute in the void volume. Molecular-weight calibration of the column suggest Sso6206 is over 100 kDa; more accurate analysis is not reliable. This suggests that Sso6206 contains in excess of ten monomers per protomer (100 kDa). As well as a peak corresponding to monomeric protein, the mass spectrum contains peaks that correspond to multimers (up to hexamers can be distinguished). Together, these data suggest that Sso6206 forms a large multimeric protomer. If such an oligomer is indeed found for the protein, it is likely to have important functional consequences. Intriguingly, neither the self-rotation function nor the native Patterson show any additional peaks which would suggest rotational or translational noncrystallographic symmetry.

The protein was targeted as part of the Scottish Structural Proteomics Facility (SSPF), which is funded by the Scottish Higher Education Funding Council (SHEFC) and the Biotechnology and Biological Sciences Research Council UK (BBSRC). JHN is a BBSRC Career Development Fellow.

References

Bowman, J. P., McCammon, S. A., Nichols, D. S., Skerratt, J. H., Rea, S. M., Nichols, P. D. & McMeekin, T. A. (1997). *Int. J. Syst. Bacteriol.* **47**, 1040–1047.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Evans, P. R. (1997). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **33**, 22–24.
Knoblauch, C., Sahm, K. & Jorgensen, B. B. (1999). *Int. J. Syst. Bacteriol.* **49**, 1631–1643.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
She, Q. *et al.* (2001). *Proc. Natl Acad. Sci. USA*, **98**, 7835–7840.