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

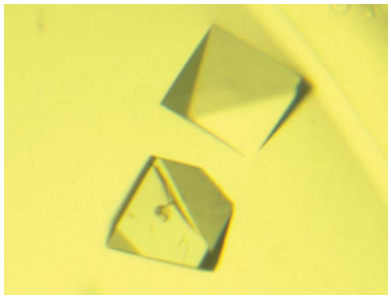
Purification, crystallization and preliminary crystallographic analysis of the non-Pfam protein AF1514 from *Archeoglobus fulgidus* DSM 4304

Pazilat Bahti,[a,b] Shunmei Chen,[b] Yang Li,[b] Neil Shaw,[b] Xuejun Zhang,[c] Min Zhang,[d] Chongyun Cheng,[b] Gaojie Song,[b] Jie Yin,[b] Hua Zhang,[e] Dongsheng Che,[e] Abdulla Abbas,[a] Hao Xu,[e] Bi-Cheng Wang[e] and Zhi-Jie Liu[b]*

[a]College of Life Science and Technology, Xinjiang University, Urumqi 830046, People's Republic of China, [b]National Laboratory of Biomacromolecules, Institution of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China, [c]Department of Immunology, Tianjin Medical University, Tianjin 300070, People's Republic of China, [d]Life Sciences College, Anhui University, Hefei 230039, People's Republic of China, and [e]SECSG, Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30605, USA

Correspondence e-mail: zjliu@ibp.ac.cn

Received 12 October 2007
Accepted 28 December 2007



A 10.5 kDa non-Pfam hypothetical protein, AF1514, from the hyperthermophilic archaeon *Archeoglobus fulgidus* has been overexpressed in *Escherichia coli*, purified and crystallized using the hanging-drop vapour-diffusion method. The crystals diffracted X-rays to 2.09 Å resolution and a data set was collected at 100 K using Cu $K\alpha$ radiation from a rotating-anode X-ray source. The crystals belong to space group $P4_12_12$ or $P4_32_12$, with unit-cell parameters $a = b = 49.27$, $c = 106.61$ Å. The calculated Matthews coefficient was 3.16 Å^3 Da^{-1}, suggesting the presence of one molecule in the asymmetric unit.

1. Introduction

The use of robotics and automation has escalated the pace of DNA sequencing and this has resulted in a number of microbial and eukaryotic genomes being sequenced completely (Bernal *et al.*, 2001). Interestingly, analysis of the genomes of each of these organisms reveals that one third of the observed open reading frames (ORFs), although conserved amongst several organisms, encode hypothetical proteins (Galperin, 2001; Galperin & Koonin, 2004). Determination of the function of these hypothetical proteins is fast emerging as a major challenge for biologists. Deducing the three-dimensional structure of these proteins could provide some clues about their function. As a result, a number of structural biology groups have prioritized their target lists to include proteins that are conserved yet have no known function (Lubec *et al.*, 2005; Eisenstein *et al.*, 2000).

Archaeoglobus fulgidus is a hyperthermophilic strictly anaerobic sulfate-reducing archaeon commonly found in marine environments (Stetter *et al.*, 1987; Stetter, 1995). Sequencing of the *A. fulgidus* genome revealed the presence of 2436 ORFs, half of which could not be annotated with any known function (Klenk *et al.*, 1997). About 30% (741) of the ORFs are non-Pfam. Non-Pfam sequences and structures are likely to confer a competitive advantage on an organism during adaptation to different environments. These unique adaptation functions could possibly be accomplished using novel protein folds. Therefore, a non-Pfam target-selection strategy will not only accelerate the expansion of protein-fold space, but will also help biologists understand the function of these proteins based on their structures (Kuratani *et al.*, 2006; Lehtio *et al.*, 2006; Chiu *et al.*, 2001; LaRonde-LeBlanc & Wlodawer, 2004). ORF AF1514 from *A. fulgidus* DSM 4304 encodes a 91-residue non-Pfam protein (GenBank accession No. NP_070343) of unknown function with a molecular weight of 10.5 kDa. A *WU-BLAST* search of the PDB for structural homologues revealed no significant matches (E value > 1). We decided to determine the three-dimensional structure of the protein by X-ray crystallography in order to gain insight into its function.

2. Materials and methods

2.1. Protein purification

The 10.5 kDa protein was expressed in *Escherichia coli* BL21 (DE3) with a TEV (tobacco etch virus protease) cleavable N-terminal hexa-His tag. Cells were grown at 310 K in LB until the

$OD_{600\,nm}$ reached 0.8, after which recombinant protein production was initiated by first cooling the culture to 285 K and then adding 0.2 m*M* IPTG. After 40 h induction, the cells were harvested by centrifugation at 4000 rev min^{-1} for 30 min and lysed by sonication. Cell debris was removed by centrifugation at 16 000*g* for 30 min and the soluble fraction was applied onto an Ni^{2+}-affinity column (Ni–NTA agarose, Qiagen) equilibrated with PBS buffer (50 m*M* potassium phosphate, 150 m*M* NaCl pH 7.2). The column was washed thoroughly with 100 ml wash buffer (PBS + 10 m*M* imidazole) and eluted with 10 ml elution buffer (PBS + 500 m*M* imidazole pH 7.2). After buffer exchange using an Amicon Ultra-4 5000 Da molecular-weight cutoff (Millipore) centrifugal concentrator, the His tag was cleaved by treating the protein with TEV for 1 h at 303 K in PBS buffer. TEV and uncut protein were removed by Ni-affinity chromatography. The protein was subjected to a final size-exclusion chromatography step using a Superdex G75 10/30 (Amersham Biosciences) column equilibrated with 20 m*M* Tris pH 8.0, 200 m*M* NaCl and 1 m*M* DTT. The protein eluted in a single peak. Fractions containing the protein were pooled and concentrated to 10–20 mg ml^{-1} prior to setting up crystallization trials. The protein was >95% pure when analyzed by SDS–PAGE.

2.2. Chemical modification

The protein was chemically modified using a reductive-methylation protocol as described previously (Shaw *et al.*, 2007). In brief, 20 µl 1 *M* dimethylamine–borane complex (DMAB) and 40 µl 1 *M* formaldehyde were added to 1 ml 10 mg ml^{-1} protein solution. The reaction mixture was incubated in the dark while shaking at 220 rev min^{-1}. The chemical additions were repeated twice at 2 h intervals. Finally, 10 µl DMAB was added and the reaction mixture was incubated overnight. Excess chemicals were removed by size-exclusion chromatography using 20 m*M* Tris pH 8.0, 200 m*M* NaCl and 1 m*M* DTT buffer solution. The methylated protein was concentrated to 10–20 mg ml^{-1} and screened for crystallization similar to the non-methylated protein.

2.3. Crystallization

Methylated and nonmethylated proteins were set up for crystallization under identical conditions using the hanging-drop vapour-diffusion method. 1 µl protein solution was mixed with an equal amount of reservoir solution and equilibrated against 300 µl reservoir solution. The initial crystallization conditions were examined using commercially available sparse-matrix screening kits from Hampton Research (Crystal Screens 1 and 2, Index and PEG/Ion Screens) and Emerald Biosystems (Wizard I and II). For both proteins, crystals appeared in a number of conditions within a week. Interestingly, the methylated protein crystallized under more conditions compared with the nonmethylated protein. In addition, the morphology of the crystals formed by the methylated protein in most crystallization conditions was better than those of the nonmethylated protein. MemFac Screen condition 1 containing 0.1 *M* sodium chloride, 0.1 *M* sodium acetate trihydrate pH 4.6 and 12%(*v*/*v*) 2-methyl-2,4-pentanediol (MPD) produced morphologically identical crystals for methylated and nonmethylated protein and was selected for further optimization. A grid screen was prepared by varying the pH from 3.5 to 5.6 in steps of 0.3 pH units and the MPD concentration was varied from 8% to 16% in steps of 2%. A mother-liquor solution containing 0.1 *M* sodium chloride, 0.1 *M* sodium acetate trihydrate pH 5.0 and 10%(*v*/*v*) MPD was found to be optimal and produced almost identical crystals in 5 d for both methylated and nonmethylated protein (Fig. 1). The crystals were further tested for X-ray diffraction.

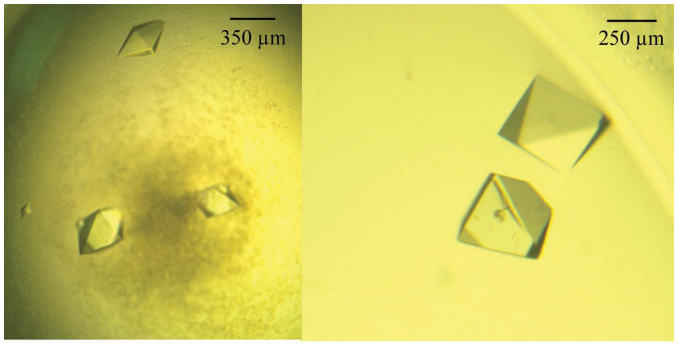


Figure 1
Nonmethylated (*a*) and methylated (*b*) protein crystals of the hypothetical protein AF1514.

2.4. X-ray diffraction studies and data collection

Crystals were harvested using a nylon loop and flash-frozen in liquid nitrogen prior to mounting. Interestingly, the crystals of the methylated protein diffracted better than the nonmethylated protein crystals. While the crystals of methylated protein diffracted X-rays to 2.07 Å resolution (Fig. 2), crystals of nonmethylated protein produced under identical conditions only diffracted to about 3.5 Å (data not shown). X-ray diffraction data were collected from a methylated protein crystal to 2.07 Å resolution at cryogenic temperature (100 K) using a Rigaku MM007 Cu $K\alpha$ rotating-anode source and a Rigaku R-AXIS IV^{++} detector. The crystal was cooled in a constant stream of nitrogen gas during data collection. A single-axis φ scan with 360 oscillation images of 1° each was recorded using a crystal-to-detector distance of 160 mm and an exposure time of 240 s per image. The data were indexed, integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). The data-collection statistics are listed in Table 1.

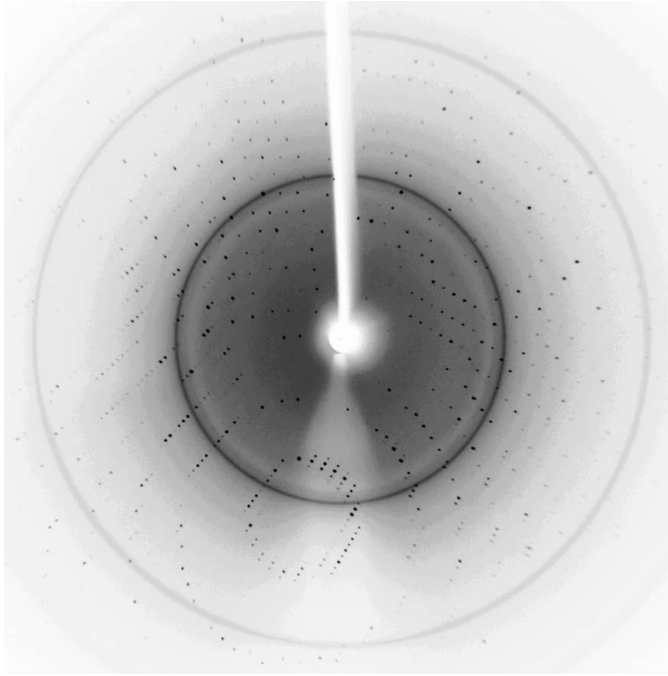


Figure 2
X-ray diffraction image from a native AF1514 crystal.

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	1.54
Space group	$P4_12_12$ or $P4_32_12$
Resolution (Å)	20.0–2.07
Total No. of reflections	92475
No. of unique reflections	8320 (653)
$I/\sigma(I)$	81.0 (39.0)
Redundancy	11.1 (11.2)
Completeness (%)	95.4 (90.0)
$R_{\rm merge}$	0.033 (0.079)
$R_{\rm r.i.m.}$†	0.046 (0.131)
$R_{\rm p.i.m.}$†	0.014 (0.039)

† Calculated according to Weiss (2001) and Evans (2006).

3. Results and discussion

The open reading frame of AF1514 consists of 273 bp encoding a 91-amino-acid protein. Methylated and nonmethylated protein could be purified to >95% purity as determined by SDS–PAGE. Both the methylated and nonmethylated proteins formed crystals under a number of conditions. Methylation of surface lysines has been shown to improve the crystallizability of proteins. The reductive-methylation protocol used in the current study usually results in the dimethylation of free amine groups. The methyl groups attached to the side-chain amine N atoms of lysine residues form cohesive (NZ)CH$\cdots$O contacts with the neighbouring electron-negative carboxyl and carbonyl O atoms. This localizes the side chain of lysine residues in space, resulting in a more compact protein molecule (Shaw *et al.*, 2007). A mother-liquor solution containing 0.1 *M* sodium acetate pH 5.0, 0.1 *M* sodium chloride, 10%(*w*/*v*) MPD produced morphologically identical crystals of both methylated and nonmethylated protein in 5 d. The crystals showed no visual defects when observed under a microscope and were suitable for X-ray diffraction studies. Since the crystallization condition contained a cryoprotectant, crystals were flash-frozen in liquid nitrogen and directly tested for X-ray diffraction. Crystals of the methylated protein that were obtained under identical conditions to those used to produce the nonmethylated protein crystals diffracted better than crystals of the nonmethylated protein. This result was not surprising because methylation affects the protein surface. Since crystallization is predominantly a surface phenomenon, a methylated protein is likely to behave differently during crystallization compared with a nonmethylated protein. Although the crystallization conditions and morphology of the crystals were similar, the packing of the molecules in the lattice may differ between the two proteins, resulting in different diffraction patterns. Crystals of the methylated protein obtained under various conditions are currently being tested for X-ray diffraction.

A native data set consisting of 8320 unique reflections was collected for the methylated protein using a copper X-ray source. The crystals were tetragonal in shape and belong to space group $P4_12_12$ or $P4_32_12$, with unit-cell parameters $a = b = 49.27$, $c = 106.60$ Å. The calculated Matthews coefficient was 3.16 Å^3 Da^{-1}, suggesting the presence of one molecule in the asymmetric unit. The crystal showed no signs of decay upon exposure to X-rays. Phasing using the anomalous signal of S atoms is currently being attempted in order to solve the structure.

This work was funded by the 863 (grant 2006AA02A316) and 973 (grant 2006CB910901) projects of the Ministry of Science and Technology of China, the National Natural Science Foundation of China (grants 30670427, 30700123 and 30770432), the CAS–KIST Collaborative Research Grant, the US National Institutes of Health (grant 1P50 GM62407), the University of Georgia Research Foundation and the Georgia Research Alliance.

References

Bernal, A., Ear, U. & Kyrpides, N. (2001). *Nucleic Acids Res.* **29**, 126–127.
Chiu, H.-J., Johnson, E., Schroder, I. & Rees, D. (2001). *Structure*, **9**, 311–319.
Eisenstein, E., Gilliland, G., Herzberg, O., Moult, J., Orban, J., Poliak, R., Banerjei, L., Richardson, D. & Howard, A. (2000). *Curr. Opin. Biotechnol.* **11**, 25–30.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Galperin, M. Y. (2001). *Funct. Genomics*, **2**, 14–18.
Galperin, M. Y. & Koonin, E. V. (2004). *Nucleic Acids Res.* **32**, 5452–5463.
Klenk, H. P. *et al.* (1997). *Nature (London)*, **390**, 364–370.
Kuratani, M. *et al.* (2006). *J. Mol. Biol.* **355**, 395–408.
LaRonde-LeBlanc, N. & Wlodawer, A. (2004). *Structure*, **12**, 1585–1594.
Lehtio, L. J., Grossmann, G., Kokona, B., Fairman, R. & Goldman, A. (2006). *J. Mol. Biol.* **357**, 221–235.
Lubec, G., Afjehi-Sadat, L., Yang, J.-W. & John, J. (2005). *Prog. Neurobiol.* **77**, 90–127.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Shaw, N., Cheng, C., Tempel, W., Chang, J., Ng, J., Wang, X.-Y., Perrett, S., Rose, J., Rao, Z., Wang, B.-C. & Liu, Z.-J. (2007). *BMC Struct. Biol.* **7**, 46–58.
Stetter, K. O. (1995). *ASM News.* **61**, 285–290.
Stetter, K. O., Lauerer, G., Thomm, M. & Neuner, A. (1987). *Science*, **236**, 822–824.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.