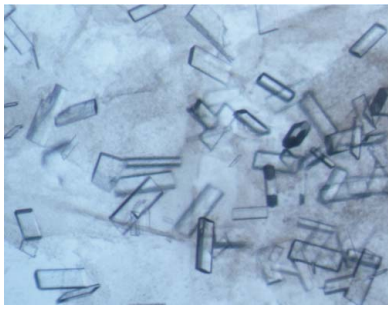
Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Byeong-Gu Han[a,b] and Byung Il Lee[a]*

[a]Cancer Cell and Molecular Biology Branch, Division of Cancer Biology, Research Institute, National Cancer Center, Goyang, Gyeonggi 410-769, Republic of Korea, and [b]Department of Biological Science, Sungkyunkwan University, Suwon, Gyeonggi 440-746, Republic of Korea

Correspondence e-mail: bilee@ncc.re.kr

Received 2 June 2011
Accepted 18 June 2011



Overexpression, crystallization and preliminary X-ray crystallographic analysis of *Pyrococcus furiosus* PF2050, a member of the DUF2666 protein family

Pyrococcus furiosus PF2050, a putative uncharacterized protein, was overexpressed in *Escherichia coli*, purified and crystallized at 298 K using 2-methyl-2,4-pentanediol as the precipitant. X-ray diffraction data were collected to a resolution of 1.56 Å at 100 K using synchrotron X-rays. The crystal was found to belong to space group $P2_1$, with unit-cell parameters a = 41.76, b = 66.43, c = 46.38 Å, β = 96.62°. The asymmetric unit contained one subunit of PF2050, with a crystal volume per protein weight (V_M) of 2.21 Å^3 Da^{-1} and a solvent content of 44.41%.

1. Introduction

Pyrococcus furiosus is an aquatic anaerobic extremophilic species of archaeon and its genome was completely sequenced in 2001 (Robb *et al.*, 2001). This has led to the identification of many functionally uncharacterized proteins. One such putative uncharacterized protein, PF2050, consists of 252 amino acids with a calculated molecular weight of 28.9 kDa. Currently, there is no information describing the function of this protein. Amino-acid sequence alignment with other proteins revealed that it contains two conserved DUF2666 domains (Schultz *et al.*, 1998; Bateman *et al.*, 2010). DUFs (domains of unknown function) comprise a large set of uncharacterized protein families that are found in the Pfam database (Finn *et al.*, 2010). Approximately three-quarters of all known proteins now match one or another of the ∼10 000 protein families in the Pfam database and the proportion of DUF families in Pfam has increased to approximately 22% of all protein families (Bateman *et al.*, 2010). Identification of the functions of DUF-family proteins is essential for understanding biological events in organisms.

Because functional or structural studies of DUF2666 domains have not yet been established, predicting their cellular roles has been difficult. *BLAST* searches using the amino-acid sequence of PF2050 show that most of the similar proteins are putative uncharacterized proteins. Only two annotated proteins, namely subunit C of the SbcCD nuclease from *Fusobacterium* sp. (amino-acid sequence identity of 24% for 230 amino acids, *E* value of 0.005) and a putative DNA helicase from *Methanococcus vannielii* (amino-acid sequence identity of 23% for 209 amino acids, *E* value of 0.011), have a low sequence identity to PF2050. Therefore, it is difficult to predict the function of this protein; instead, structural information for PF2050 will be required in order to provide information on the function of the protein. Therefore, as a first step towards elucidating the PF2050 structure, we overexpressed, purified and crystallized the protein for structural determination.

2. Materials and methods

2.1. Overexpression and purification

The *P. furiosus PF2050* gene was amplified from genomic DNA by polymerase chain reaction (PCR) and cloned into a modified pET32 vector (Novagen) using the *Nde*I and *Xho*I restriction sites. The protein was overexpressed in *Escherichia coli* Rosetta2 (DE3) cells (Novagen). The cells were grown at 310 K in 4 l Terrific Broth medium to an OD_{600} of 0.7 and expression of the recombinant protein was induced by 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside

Table 1
Sample information.

Macromolecule details	
Component molecules	PF2050
Macromolecular assembly	Monomer
Mass (Da)	28900
Source organism	*E. coli* strain Rosetta2 (DE3)
Crystallization and crystal data	
Crystallization method	Hanging-drop vapour diffusion
Temperature (K)	298
Apparatus	24-well plates
Atmosphere	Room air
Crystal-growth time (d)	3
Crystallization solutions	
Macromolecule and storage buffer	2 µl PF2050 (16 mg ml^{-1}), 20 m*M* Tris–HCl pH 7.5, 200 m*M* NaCl and 1 m*M* DTT
Screen solution	2 µl 20 m*M* $CaCl_2$, 100 m*M* sodium acetate trihydrate pH 4.6 and 30%(*v*/*v*) MPD
Reservoir	500 µl 20 m*M* $CaCl_2$, 100 m*M* sodium acetate trihydrate pH 4.6 and 30%(*v*/*v*) MPD
Crystal data	
Crystal dimensions (mm)	0.5 × 0.2 × 0.1
Matthews coefficient V_M (Å^3 Da^{-1})	2.21
Solvent content (%)	44.41
Unit-cell data	
Crystal system, space group	Monoclinic, $P2_1$
Unit-cell parameters (Å, °)	a = 41.76, b = 66.43, c = 46.38, $\alpha = \gamma = 90$, $\beta = 96.62$

(IPTG) at 303 K. Cell growth continued at 303 K for 16 h after IPTG induction and the cells were harvested by centrifugation at 3000*g* for 10 min at 277 K. The cell pellet was suspended in ice-cold lysis buffer [25 m*M* Tris–HCl pH 7.4, 138 m*M* NaCl, 2 m*M* KCl, 10%(*v*/*v*) glycerol and 1 m*M* phenylmethylsulfonyl fluoride] containing 0.8 m*M* lysozyme and homogenized by sonication. The crude lysate was centrifuged at 39 191*g* for 1 h at 277 K and the cell debris was discarded. The first purification step was heat treatment. The supernatant was heated to 358 K for 10 min and the denatured proteins were removed by centrifugation. Further purification was achieved by anion-exchange chromatography on a Q Sepharose Fast Flow column (GE Healthcare) previously equilibrated with buffer consisting of 25 m*M* Tris–HCl pH 7.4, 138 m*M* NaCl, 2 m*M* KCl and 10%(*v*/*v*) glycerol. Rather than binding and eluting the PF2050 from the column resin, the anion exchanger was used to bind impurities; PF2050 was allowed to flow through the column. The final purification step was gel filtration on a HiLoad 16/60 Superdex 75 prep-grade column (GE Healthcare) previously equilibrated with buffer consisting of 20 m*M* Tris–HCl pH 7.5, 200 m*M* NaCl and 1 m*M* DTT.

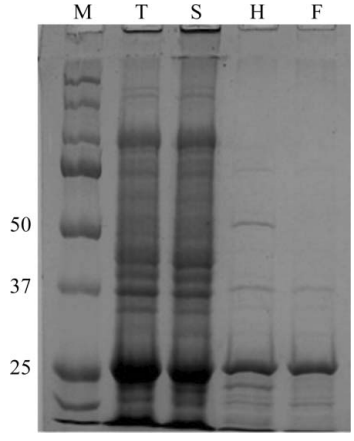
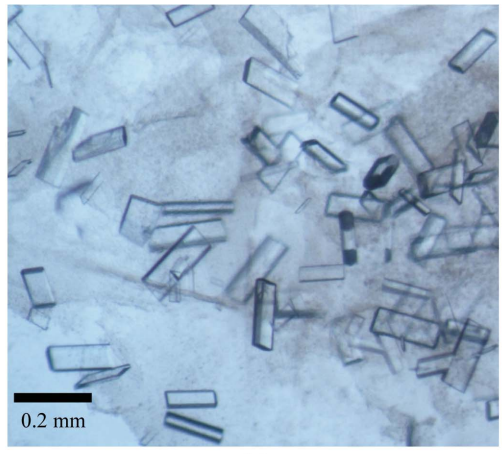


Figure 1
SDS–PAGE analysis of each purification step. Lane M, protein marker (labelled in kDa); lane T, protein expression; lane S, supernatant after cell lysis and centrifugation; lane H, supernatant after boiling and centrifugation; lane F, flowthrough sample from the ion-exchange column.

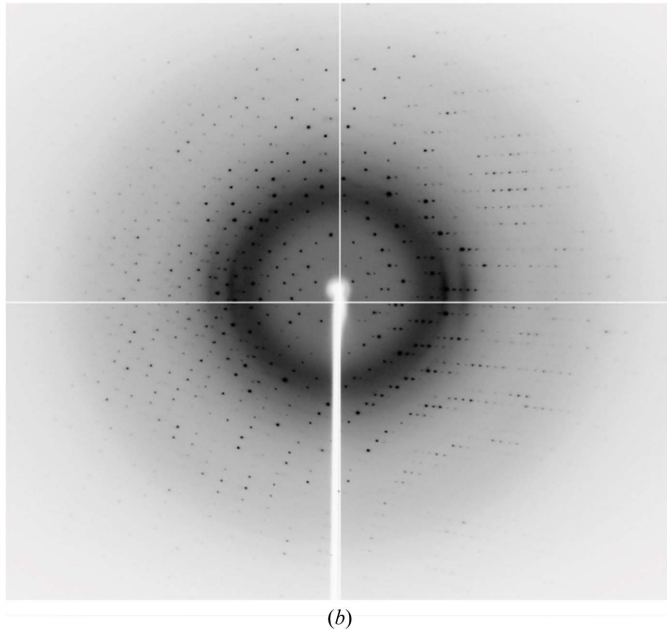
The protein purity was analyzed by SDS–PAGE (Fig. 1). The purified protein solution was concentrated to 16 mg ml^{-1} using a YM10 membrane (Millipore). The protein concentration was estimated by measuring the absorbance at 280 nm, employing the molar extinction coefficient of 19 940 M^{-1} cm^{-1} (0.690 mg ml^{-1}) calculated using the ExPASy Proteomics Server (http://www.expasy.org/).

2.2. Crystallization and X-ray data collection

Initial crystallization conditions were investigated employing the sitting-drop crystallization method using 96-well plates (Axygen Biosciences) at 298 K. The crystallization drop was prepared by mixing 0.7 µl protein solution and the same amount of reservoir solution. The Crystal Screen, Crystal Screen 2 and MembFac kits (Hampton Research) were employed. Initial crystallization conditions were further optimized by the hanging-drop vapour-diffusion method. Each hanging drop was prepared by mixing 2 µl protein solution and the same amount of reservoir solution and was placed over 500 µl reservoir solution. For data collection, the crystals were mounted on nylon loops, washed with reservoir solution and subsequently flash-cooled in liquid nitrogen. Since the reservoir solution



(*a*)



(*b*)

Figure 2
(*a*) Crystals of PF2050. (*b*) Diffraction pattern of the PF2050 crystal.

Table 2
Data-collection statistics.

Values in parentheses are for the outer shell.

Diffraction source	PF 5A
Wavelength (Å)	1.0000
Detector	ADSC Quantum 210 CCD
Temperature (K)	100
Resolution range (Å)	50–1.56 (1.59–1.56)
No. of unique reflections	34528
No. of observed reflections	179754
Completeness (%)	95.8 (88.2)
Multiplicity	5.2 (4.5)
$\langle I/\sigma(I)\rangle$	40.6 (6.2)
$R_{\rm merge}$†	0.071 (0.304)
Data-processing software	*HKL*-2000

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of reflection hkl and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all i measurements.

acted as a cryoprotectant, no search for a separate cryoprotectant solution was necessary. X-ray diffraction data were collected at 100 K using an ADSC Quantum 210 CCD detector at the 5A experimental station of the Photon Factory (PF), Japan. The crystals were rotated through a total of 270° with a 1.0° oscillation range per frame. The raw data were processed and scaled using *HKL*-2000 (Otwinowski & Minor, 1997).

3. Results and discussion

The recombinant PF2050 protein was overexpressed in *E. coli* in a soluble form and purified to give a yield of ∼10 mg purified protein per litre of culture. The homogeneity of the purified protein and additional sample information are shown in Fig. 1 and Table 1, respectively. Crystals of PF2050 were grown in several crystallization conditions. The best PF2050 crystal was obtained using a reservoir solution consisting of 0.02 *M* calcium chloride dehydrate, 0.1 *M* sodium acetate trihydrate pH 4.6 and 30% 2-methyl-2,4-pentanediol. Crystals grew to suitable sizes for X-ray diffraction experiments within 3 d (Fig. 2). The approximate dimensions of the crystals were 0.5 × 0.2 × 0.1 mm (Fig. 2). The crystals diffracted to at least 1.5 Å resolution using synchrotron radiation and diffraction data were collected to 1.56 Å (Fig. 2). A total of 179 754 measured reflections were merged into 34 528 unique reflections with an $R_{\rm merge}$ of 7.1%. The merged data set was 95.8% complete to 1.56 Å. The systematic absences indicated that the space group was $P2_1$, with unit-cell parameters a = 41.76, b = 66.43, c = 46.38 Å, β = 96.62°. The presence of one subunit of PF2050 in an asymmetric unit gave a crystal volume per protein mass ($V_{\rm M}$) of 2.21 Å^3 Da^{-1} and a corresponding solvent content of 44.41% (Matthews, 1968). The statistics for data collection are summarized in Tables 1 and 2. A three-wavelength MAD data set has also been collected from a selenomethionine-substituted protein crystal in order to determine the phases using the MAD method.

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (grant No. 2010-0020993 to BIL).

References

Bateman, A., Coggill, P. & Finn, R. D. (2010). *Acta Cryst.* F**66**, 1148–1152.
Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R. & Bateman, A. (2010). *Nucleic Acids Res.* **38**, 211–222.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Robb, F. T., Maeder, D. L., Brown, J. R., DiRuggiero, J., Stump, M. D., Yeh, R. K., Weiss, R. B. & Dunn, D. M. (2001). *Methods Enzymol.* **330**, 134–157.
Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 5857–5864.