FEBS Letters 586 (2012) 1384–1388

journal homepage: www.FEBSLetters.org

Crystal structure of *Pyrococcus furiosus* PF2050, a member of the DUF2666 protein family

Byeong-Gu Han [a,d], Kyung-Chae Jeong [a], Jea-Won Cho [a], Byung-Cheon Jeong [b], Hyun Kyu Song [b], Jae Young Lee [c], Dong Hyuk Shin [d], Sangho Lee [d], Byung Il Lee [a,*]

[a] Biomolecular Function Research Branch, Division of Convergence Technology, Research Institute, National Cancer Center, Goyang, Gyeonggi 410-769, Republic of Korea
[b] School of Life Sciences and Biotechnology, Korea University, Seoul 136-701, Republic of Korea
[c] Department of Life Science, Dongguk University, Seoul 100-715, Republic of Korea
[d] Department of Biological Science, Sungkyunkwan University, Suwon, Gyeonggi 440-746, Republic of Korea

ARTICLE INFO

Article history:
Received 26 December 2011
Revised 28 March 2012
Accepted 3 April 2012
Available online 13 April 2012

Edited by Christian Griesinger

Keywords:
PF2050
DUF2666
DNA binding
Pyrococcus furiosus

ABSTRACT

Pyrococcus furiosus **PF2050 is an uncharacterized putative protein that contains two DUF2666 domains. Functional and structural studies of PF2050 have not previously been performed. In this study, we determined the crystal structure of PF2050. The structure of PF2050 showed that the two DUF2666 domains interact tightly, forming a globular structure. Each DUF2666 domain comprises an antiparallel β-sheet and an α-helical bundle. One side of the PF2050 structure has a positively charged basic cleft, which may have a DNA-binding function. Furthermore, we confirmed that PF2050 interacts with circular and linear dsDNA.**

1. Introduction

The DUFs (Domain of Unknown Function) are a large set of uncharacterized protein families that are found in the Pfam database [1]. To fully understand the biological processes that sustain life, identification of the functions of the DUF family proteins is essential [2]. Structural determination can be a good strategy for functional identification. To date, the three-dimensional structures of some DUFs have been determined, and their functional roles have been revealed. For example, TM841 (DUF194) from *Thermotoga maritima* was found to be a fatty acid-binding protein [3], NE1406 (DUF2006) from *Nitrosomonas europaea* belongs to the calycin superfamily [4], and the DUF283 domain of DCL4 from *Arabidopsis thaliana* has a dsRNA-binding fold that is implicated in protein–protein interactions [5].

The PF2050 protein from the thermophilic archaeon *Pyrococcus furiosus* contains two DUF2666 domains [6]. The function of DUF2666 has not yet been determined. When we investigated the *P. furiosus* genome, several genes that regulate transcription were identified near *PF2050*, suggesting their functional role in the transcriptional regulation of *PF2050*. Among the neighboring loci, *PF2053* (AsnC family) and *PF2051* (ArsR family) are transcription regulator proteins. These genes are moderately conserved among *P. furiosus*, *Pyrococcus abyssi*, *Pyrococcus horikoshii*, and *Thermococcus kodakaraensis*. Various proteins homologous to PF2050 have been identified by a BLAST search [7]. In contrast to the *P. abyssi* and *Thermococcus gammatolerans* homologs that each have two DUF2666 domains, the *Methanococcus jannaschii* homolog has only one DUF2666 domain.

Structural and functional studies of PF2050 are necessary to characterize the function of the DUF2666 family of proteins because the function of the DUF2666 domain has not been elucidated yet. Therefore, as a first step toward the functional characterization of the DUF2666 protein family, we determined the crystal structure of PF2050. The DNA-binding function of PF2050 was also demonstrated.

2. Materials and methods

2.1. Crystallization and data collection

The recombinant protein was expressed, purified and crystallized as described elsewhere [8]. Crystals were 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. The crystal belonged to the $P2_1$ space group with unit-cell parameters of $a = 41.72$ Å, $b = 66.53$ Å, $c = 46.37$ Å,

* Corresponding author.
E-mail address: bilee@ncc.re.kr (B.I. Lee).

0014-5793/$36.00
http://dx.doi.org/10.1016/j.febslet.2012.04.004

$\alpha = \gamma = 90°$, and $\beta = 96.56°$. Native data were collected at a resolution of 1.56 Å. The asymmetric unit contains one monomer (Table 1).

2.2. Structure determination and refinement

One set of MAD data from the selenomethionine-substituted (SeMet) protein crystal was used for phase calculations. The selenium sites were located with SOLVE [9], and the initial phases were improved using the program RESOLVE [9]. The resulting electron density map was interpretable for model building. Manual model building was performed using the program COOT [10], and the model was refined with the program REFMAC5 [11]. Five percent of the data was randomly set aside for use as the test data for the calculation of R_{free}. The refined model has excellent stereochemistry, which was evaluated with the program MolProbity [12].

2.3. In vitro DNA-binding assays

Purified recombinant PF2050 was concentrated to 5 mg/ml in buffer (20 mM Tris–HCl pH 7.5, 200 mM NaCl, and 1 mM DTT). The p3XFLAG plasmid (6299 bp, Sigma) was purified for the DNA-binding assay. Linearized dsDNA was made by cutting the plasmid DNA with *Eco*RI. Various PCR products were also used for experiments with the linearized dsDNA. Plasmid DNA (12.8 nM) was mixed with purified protein (3.46, 6.92, 17.1, or 34.6 μM) in a reaction buffer (100 mM Tris–HCl pH 8.0). Total reaction volume was 50 μl. Reaction mixtures were incubated at 4 °C for one hour. Various plasmids (pET32, pcDNA 3.0 and pAcHLT-C) and PCR products (HIST1H4A, API5, and NCAPG2) were also tested. DNA binding was analyzed by 1% agarose gel electrophoresis in TAE buffer. To test DNA-binding activity at low protein concentration, radioisotope labeled dsDNA was used for DNA-binding assay. PCR products (HIST1H4A: 334 base pairs, PYY peptide: 138 base pairs) were labeled using [γ-^{32}P] ATP (Izotop, Hungary) and T4 polynucleotide kinase (TaKaRa bio, Japan). The unlabeled [γ-^{32}P] ATP was removed with Sephadex G-25 column (GE healthcare, US) according to the manufacturer's instructions. The labeled oligonucleotides (10 nM) were incubated with each concentration of PF2050 protein (1 μM, 100 nM, and 10 nM) in a reaction buffer (10 mM Tris–HCl pH 7.0, 50 mM NaCl, 1 mM $MgCl_2$, 0.5 mM EDTA, 0.5 mM DTT, and 4% glycerol). Total reaction volume was 20 μl. In order to achieve a state of equilibrium, we incubated the whole mixtures at RT for 20 min. The protein–DNA complexes were separated from the free DNA on a 6% polyacrylamide gel prepared and pre-electrophoresed in TBE buffer. Electrophoresis was performed in the TBE buffer at 100 V and RT for 1 h. Each band was visualized by autoradiography.

2.4. Data deposition

The atomic coordinates and structure factors have been deposited in the Protein Data Bank (http://www.rcsb.org/pdb/home/home.do) with the accession number 3V68.

3. Results and discussion

3.1. Overall and DUF2666 domain structure

We determined the crystal structure of PF2050 at a resolution of 1.56 Å. The refined model exhibited working and free *R* values of 19.8% and 22.7%, respectively, for the 50.0–1.56 Å data without a sigma cutoff. Crystallographic data, phasing, and refinement statistics are shown in Table 1. The model includes four β-strands, 10 α-helices, and 248 residues for the PF2050 monomer (Fig. 1A). Only the first methionine of the N-terminus is missing in the model. The two DUF2666 domains of PF2050 form a closed clam-like structure. The two DUF2666 domains interact tightly to form a globular structure (Fig. 1A). One side of the structure is comprised of four β-strands, and the other side of the structure contains 10 α-helices. The structure can be divided into two regions, the DUF2666-1 domain (residues 2–134) and the DUF2666-2 domain (residues 135–249), which represent the two DUF2666 domains (Fig. 1A and C). Although the two DUF2666 domains in PF2050 have only approximately 15.7% amino acid sequence identity, the overall folds of the two DUF2666 domains are generally similar. In particular, the N-terminal regions of the DUF2666 domains, β1–β2–α1 residues 2–47 and β3–β4–α6 residues 135–181, are quite similar to each other. The root-mean-square deviation (r.m.s.d.) between the β1–β2–α1 and β3–β4–α6 regions is 2.23 Å for 44 C_α atom pairs, and the sequence identity for the two regions is 27.3% (Fig. 1B and C). The three helices at the C-terminal region of the DUF2666 domains are folded into simple helical bundles (α2, α3, and α5 in DUF2666-1 and α7, α8, and α9 in DUF2666-2,

Table 1
Data collection and refinement statistics.

Crystal	SeMet			Native
Data collection				
X-ray source	PF-NW12A			PF-5A
Space group	$P2_1$ (native cell parameters, a = 41.72 Å, b = 66.53 Å, c = 46.37 Å, and β = 96.56°)			
Data set	SeMet (peak)	SeMet (edge)	SeMet (Remote)	Native
Wavelength (Å)	0.97910	0.97924	0.96000	1.00000
Resolution (Å)[a]	50–1.6 (1.63–1.60)	50–1.6 (1.63–1.60)	50–1.6 (1.63–1.60)	50–1.56 (1.59–1.56)
Redundancy[a]	5.1 (5.1)	5.1 (5.1)	5.1 (5.2)	7.4 (6.3)
Completeness (%)[a]	95.4 (93.4)	95.4 (93.7)	95.5 (93.9)	94.8 (74.3)
I/σ_I[a]	39.3 (4.6)	38.6 (4.3)	37.9 (4.1)	61.3 (9.7)
R_{merge} (%)[a]	9.0 (44.3)	9.1 (47.4)	8.9 (48.1)	3.9 (23.9)
Phasing				
No. of Se sites	5			
Figure of merit	0.52/0.65 (before/after RESOLVE)			
Refinement				
Resolution (Å)				50.0–1.56
No. of reflections				32,516
R_{work}/R_{free} (%)				19.8/22.7
No. of atoms				2289
Average B-factors	(protein/water/MPD)			17.8/32.0/24.4
R.m.s. deviations				0.006/0.960

[a] Values in the parentheses refer to the highest resolution shells.