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

ARTICLE *in* FEBS LETTERS · MAY 2012

Impact Factor: 3.17 · DOI: 10.1016/j.febslet.2012.04.004 · Source: PubMed

9 AUTHORS, INCLUDING:

Byeong-Gu Han

14 PUBLICATIONS **142** CITATIONS

SEE PROFILE

Byung-Cheon Jeong

University of Texas Southwestern Medical …

17 PUBLICATIONS **120** CITATIONS

SEE PROFILE

Hyun Kyu Song

Korea University

110 PUBLICATIONS **3,100** CITATIONS

SEE PROFILE

Donghyuk Shin

Sungkyunkwan University

12 PUBLICATIONS **20** CITATIONS

SEE PROFILE

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.

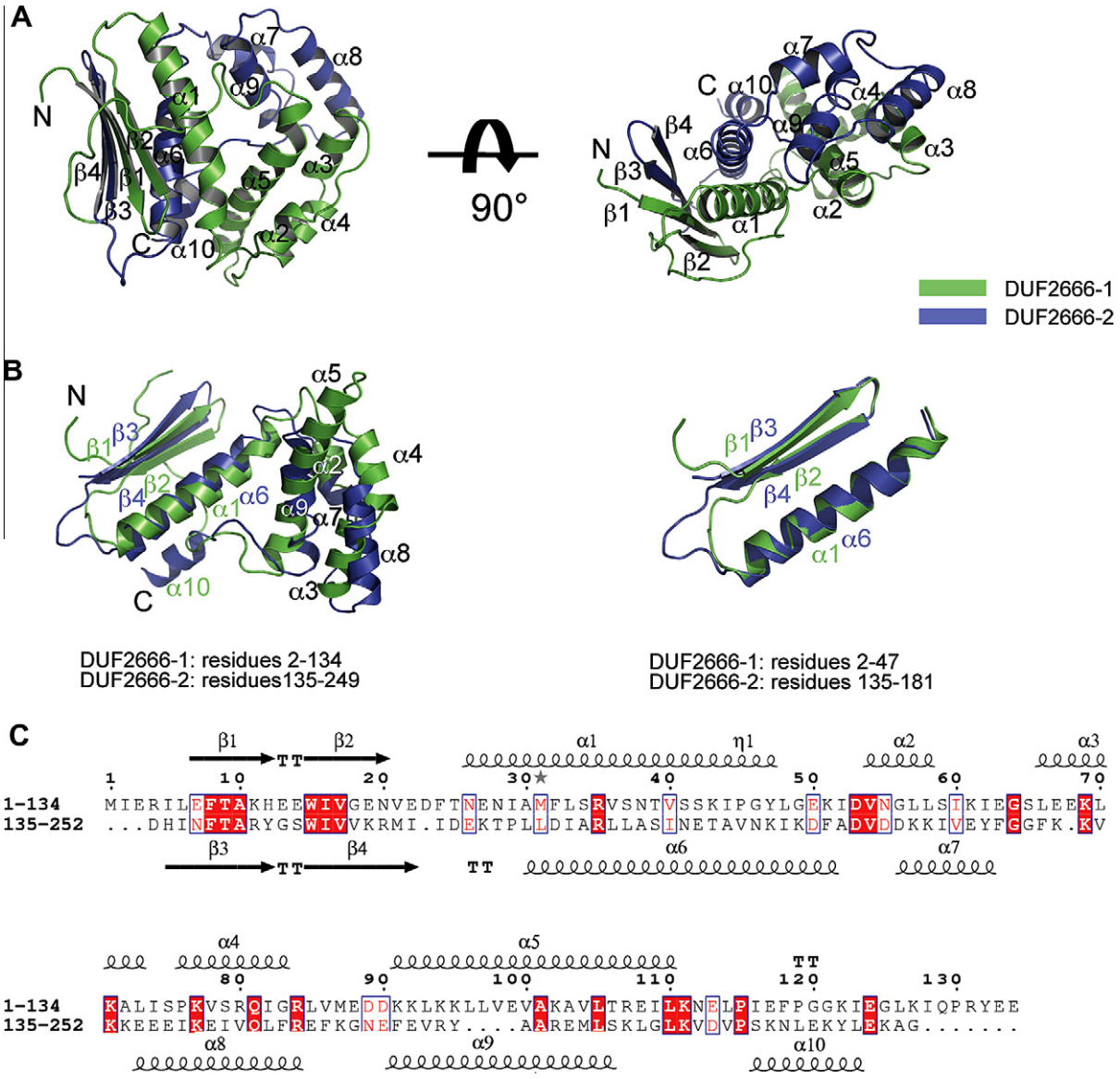

Fig. 1. The overall structure of PF2050 and the structural similarities of the DUF2666 domains. (A) The monomeric structure of PF2050 contains four β-strands and 10 α-helices. The DUF2666 domains are drawn in green and blue. (B) DUF2666–1 is structurally similar to DUF2666-2 (left). β1–β2–α1 is structurally similar to β3–β4–α6 (right). DUF2666-1 and DUF2666-2 are drawn in green and blue, respectively. (C) The amino acid sequence alignment between the two DUF2666 domains of PF2050 is shown.

and their overall folds are similar. However, the structures of these helix bundles deviate more from each other than the N-terminal regions of two DUF2666 domains (Fig. 1B) do. Helix α10 only exists in the DUF2666-2 domain (Fig. 1B and C).

The two DUF2666 domains tightly interact via hydrogen bonding and hydrophobic interactions. The N-terminal region of each domain starts with a β-strand–β-hairpin–β-strand motif, and β1 from the DUF2666-1 domain and β3 from the DUF2666-2 domain interact strongly, forming an antiparallel β-sheet (Fig. 2A). Eleven hydrogen bonds (8 through the main chain and 3 through the side chains) were found between the β1 and β3 strands (Fig. 2A). Additionally, there are intradomain hydrogen bonds between β1 and β2 and between β3 and β4 that stabilize the DUF2666 domains themselves (Fig. 2A). Interactions via antiparallel β-sheets are known to be a fundamental mode of protein–protein interactions that are important for protein-structure stabilization [13]. Such antiparallel β-sheet structures have also been found in other dimeric proteins, such as an artificial β-sheet dimer, defensin HNP-3, and interleukin 8 [13–15]. Additional important interactions between the two DUF2666 domains are hydrogen bond and hydrophobic interactions between helices α1 and α6. One hydrogen bond between the side chains of Ser37 and Asn171 and many hydrophobic interactions between α1 (Ile29, Leu33, Val36, Val40, Ile44, and Tyr47) and α6 (Phe181, Ile178, Ala174, Ile170, Leu167, and Ile163) were observed (Fig. 2B). When the two DUF2666 domains were separately expressed in *Escherichia coli*, the recombinant proteins were insoluble, which could indicate the importance of the interdomain interactions for protein stability.

DALI searches were performed to identify structurally similar proteins [16]. However, no significantly similar structures were found. When DALI searches were performed separately on the DUF2666-1 domain and the DUF2666-2 domain, some protein structures were structurally aligned with a low *Z*-score. DUF2666-1 structurally aligned with the human α-hemoglobin stabilizing protein (AHSP) (PDB code: 1Y01, *Z*-score = 5.3, r.m.s.d. = 2.9 Å for 62 structurally aligned residues), which is an erythroid-specific molecular chaperone that enhances α-hemoglobin stability [17]. DUF2666-2 aligned with *Streptococcus mutans* polyribonucleotide

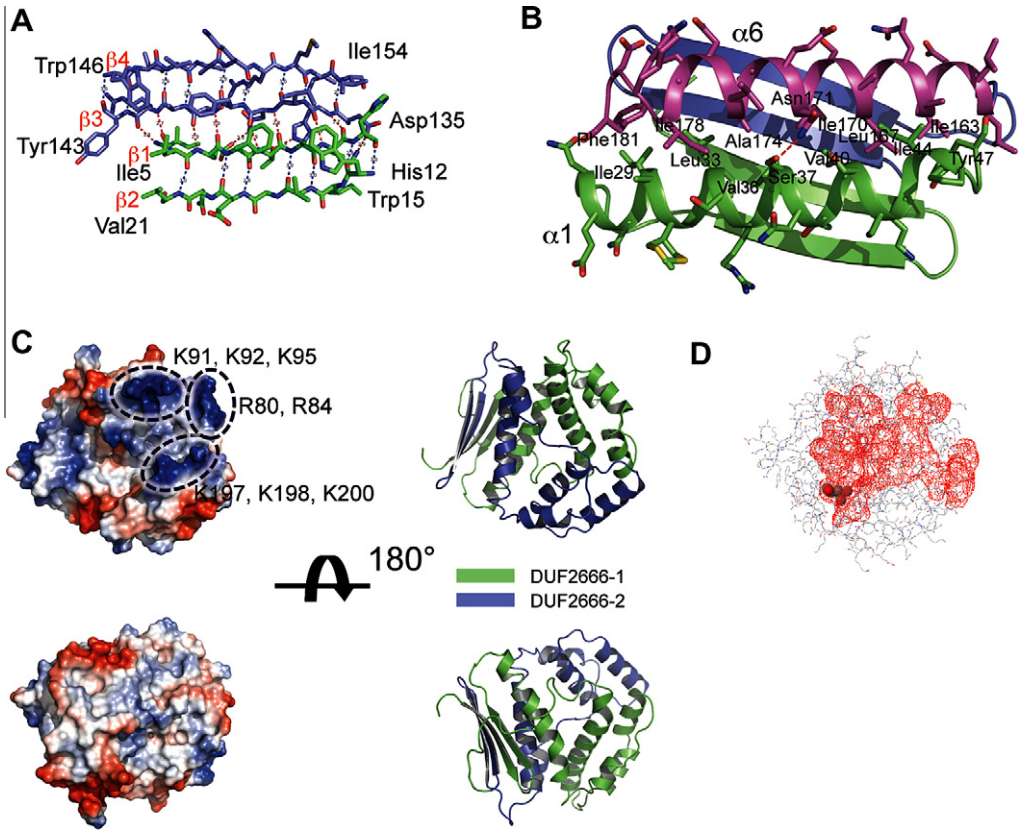

Fig. 2. The detailed view of PF2050 structure. (A) Antiparallel β-sheets interactions in the N-terminal regions of the DUF2666 domains are shown. The dotted lines indicate the hydrogen bonds between the β-strands. (B) The binding interface between α1 (green) and α6 (purple) is depicted. Ser37 and Asn171 form a hydrogen bond, and other residues are involved in hydrophobic interactions. (C) The surface charge representation of PF2050 is indicated. Positively charged regions are shown in the circles. (D) The largest cleft in the surface of the protein is shown. The red-colored region represents the positively charged cleft that can bind DNA. The sphere indicates the MPD molecule.

nucleotidyltransferase (PDB code: 3H36, *Z*-score = 4.7, r.m.s.d. = 2.6 Å for 48 structurally aligned residues), which is annotated as an RNA-binding protein that is involved in mRNA degradation [18]. However, the relatively low *Z*-scores and the short alignment lengths suggest that PF2050 has unique structural features. When a DALI search was performed with the first 49 amino acids of the N-terminus (2–50) that correspond to β1–β2–α1 of the DUF2666-1 domain, the de novo protein M7 structurally aligned to the β1–β2–α1 motif (PDB code: 2JVF, *Z*-score = 3.5, r.m.s.d. = 2.4 Å for 39 structurally aligned residues). The de novo protein M7 is a designed protein for the study of cooperative folding and has a high thermodynamic stability [19]. Therefore, it is possible that the β1–β2–α1 motif of DUF2666 is important for cooperative folding and thermodynamic stability. Although the β1–β2–α1 and β3–β4–α6 motifs are structurally similar to each other, a DALI search with the β3–β4–α6 motif of DUF2666-2 did not yield any structurally aligned proteins with reasonable *Z*-score and r.m.s.d values.

3.2. DNA-binding property

Although the BLAST searches yielded predominantly uncharacterized putative proteins as PF2050 homologs, two annotated proteins, nuclease sbcCD subunit C from *Fusobacterium* spp. and the putative DNA helicase from *Methanococcus vannielii*, partially aligned with PF2050 with a low sequence identity. These proteins are known to bind DNA. Furthermore, the surface-charge representation of PF2050 shows that one side of the protein is highly positive (Fig. 2C). Structural analysis using the PDBsum (http://www.ebi.ac.uk/pdbsum/) also shows that the positively charged region of the protein forms a large cleft with a volume of 4279.08 Å^3 and an average depth of 13.35 Å (Fig. 2D). This large cleft is of sufficient size to bind macromolecules, such as DNA or protein. The DP-Bind server (a web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins) [20] also showed that PF2050 may be a DNA-binding protein. Taken together, we assumed that PF2050 could have a DNA-binding function. To test the DNA-binding function of PF2050, we performed a DNA-binding assay. We found that PF2050 binds to dsDNA in a dose-dependent manner (Fig. 3A and B). PF2050 bound to various plasmids and linearized DNA (Fig. S2). We could not determine in this study if DNA-binding is DNA sequence specific or not. Further studies on DNA sequence specificity are required for better understanding its cellular function. The highly conserved basic amino acids, such as Arg80, Lys91, Lys92, Lys94, Lys95, Lys197, Lys198, and Lys200, were clustered in the cleft (Figs. 2C and S1). This large positively charged cleft is expected to be an important region for the DNA-binding properties of PF2050 (Fig. 2C).

YbaB (DUF149) protein shows a similar story of revealing the DNA-binding property of the functionally uncharacterized DUF protein. The three-dimensional structure of the YbaB from *Haemophilus influenzae* had shown that it exhibits a novel tweezer-like structure, suggesting the DNA binding activity [21]. And then, its DNA-binding function was confirmed by the following-up study [22].

In conclusion, the first crystal structure of a DUF2666-domain-containing protein, PF2050, was determined. The overall fold of the DUF2666 domain was unique, and the N-terminal β1–β2–α1 motif of the DUF2666 domain seems to be important for protein folding and thermodynamic stability. The positively charged large cleft of PF2050 may have a DNA-binding function.

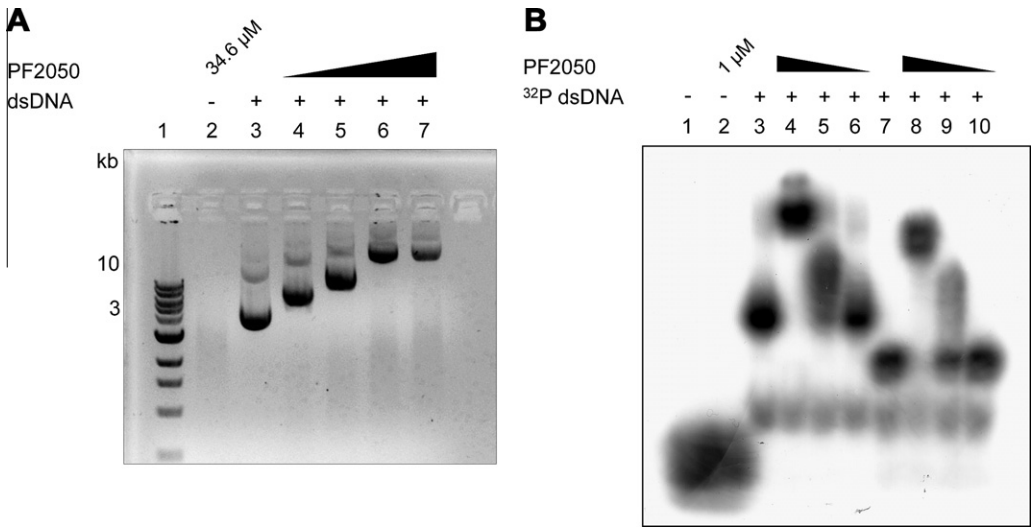

Fig. 3. The DNA-binding of PF2050. (A) PF2050 binds to circular dsDNA in a dose-dependent manner. Lane 1, DNA ladder; lane 2, protein (34.6 μM); lane 3, plasmid (p3XFLAG) DNA (12.8 nM); lane 4, plasmid (12.8 nM) and protein (3.46 μM); lane 5, plasmid (12.8 nM) and protein (6.92 μM); lane 6, plasmid (12.8 nM) and protein (17.1 μM); and lane 7, plasmid (12.8 nM) and protein (34.6 μM). (B) PF2050 binds to linear dsDNA. Lane 1, [γ-^{32}P] ATP; lane 2, [γ-^{32}P] ATP and protein (1 μM); lane 3, dsDNA (10 nM); lane 4, dsDNA (10 nM) and protein (1 μM); lane 5, dsDNA (10 nM) and protein (100 nM); lane 6, dsDNA (10 nM) and protein (10 nM); lane 7, dsDNA (10 nM); lane 8, dsDNA (10 nM) and protein (1 μM); lane 9, dsDNA (10 nM) and protein (100 nM); lane 10, dsDNA (10 nM) and protein (10 nM).

Acknowledgments

We thank the staff at the Photon Factory (Japan) beamlines 5A and NW12A for assistance with the collection of the synchrotron data. This work was supported by the Basic Science Research Program (2009-0064219 and 2009-0074586), the Mid-career Researcher Program (2011-0029294) and the Bio & Medical Technology Development Program (2011-0030032) through the National Research Foundation of Korea (NRF), and was funded by the Ministry of Education, Science and Technology to S.L. and B.I.L.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.febslet.2012.04.004.

References

[1] 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. and Bateman, A. (2010) The Pfam protein families database. Nucleic Acids Res. 38, 211–222.
[2] Bateman, A., Coggill, P. and Finn, R.D. (2010) DUFs: families in search of function. Acta Crystallogr. F Struct. Biol. Cryst. Commun. 66, 1148–1152.
[3] Schulze-Gahmen, U., Pelaschier, J., Yokota, H., Kim, R. and Kim, S.H. (2003) Crystal structure of a hypothetical protein, TM841 of *Thermotoga maritima*, reveals its function as a fatty acid-binding protein. Proteins 50, 526–530.
[4] Chiu, H.J., Bakolitsa, C., Skerra, A., Lomize, A., Carlton, D., Miller, M.D., Krishna, S.S., Abdubek, P., Astakhova, T., Axelrod, H.L., Clayton, T., Deller, M.C., Duan, L., Feuerhelm, J., Grant, J.C., Grzechnik, S.K., Han, G.W., Jaroszewski, L., Jin, K.K., Klock, H.E., Knuth, M.W., Kozbial, P., Kumar, A., Marciano, D., McMullan, D., Morse, A.T., Nigoghossian, E., Okach, L., Paulsen, J., Reyes, R., Rife, C.L., van den Bedem, H., Weekes, D., Xu, Q., Hodgson, K.O., Wooley, J., Elsliger, M.A., Deacon, A.M., Godzik, A., Lesley, S.A. and Wilson, I.A. (2010) Structure of the first representative of Pfam family PF09410 (DUF2006) reveals a structural signature of the calycin superfamily that suggests a role in lipid metabolism. Acta Crystallogr. F Struct. Biol. Cryst. Commun. 66, 1153–1159.
[5] Qin, H., Chen, F., Huan, X., Machida, S., Song, J. and Yuan, Y.A. (2010) Structure of the *Arabidopsis thaliana* DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein–protein interaction. RNA 16, 474–481.
[6] Robb, F.T., Maeder, D.L., Brown, J.R., DiRuggiero, J., Stump, M.D., Yeh, R.K., Weiss, R.B. and Dunn, D.M. (2001) Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology. Methods Enzymol. 330, 134–157.
[7] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
[8] Han, B.G. and Lee, B.I. (2011) Overexpression, crystallization and preliminary X-ray crystallographic analysis of *Pyrococcus furiosus* PF2050, a member of the DUF 2666 protein family. Acta Crystallogr. F Struct. Biol. Cryst. Commun. 67, 930–932.
[9] Terwilliger, T.C. and Berendzen, J. (1999) Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. 55, 849–861.
[10] Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132.
[11] Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. 53, 240–255.
[12] Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S. and Richardson, D.C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 35, 375–383.
[13] Khakshoor, O., Lin, A.J., Korman, T.P., Sawaya, M.R., Tsai, S.C., Eisenberg, D. and Nowick, J.S. (2010) X-ray crystallographic structure of an artificial beta-sheet dimer. J. Am. Chem. Soc. 132, 11622–11628.
[14] Hill, C.P., Yee, J., Selsted, M.E. and Eisenberg, D. (1991) Crystal structure of defensin HNP-3, an amphiphilic dimer: mechanisms of membrane permeabilization. Science 251, 1481–1485.
[15] Clore, G.M., Appella, E., Yamada, M., Matsushima, K. and Gronenborn, A.M. (1990) Three-dimensional structure of interleukin 8 in solution. Biochemistry 29, 1689–1696.
[16] Holm, L. and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res. 38, 545–549.
[17] Feng, L., Gell, D.A., Zhou, S., Gu, L., Kong, Y., Li, J., Hu, M., Yan, N., Lee, C., Rich, A.M., Armstrong, R.S., Lay, P.A., Gow, A.J., Weiss, M.J., Mackay, J.P. and Shi, Y. (2004) Molecular mechanism of AHSP-mediated stabilization of alpha-hemoglobin. Cell 119, 629–640.
[18] Ajdic, D., McShan, W.M., McLaughlin, R.E., Savic, G., Chang, J., Carson, M.B., Primeaux, C., Tian, R., Kenton, S., Jia, H., Lin, S., Qian, Y., Li, S., Zhu, H., Najar, F., Lai, H., White, J., Roe, B.A. and Ferretti, J.J. (2002) Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. Proc. Natl. Acad. Sci. U S A 99, 14434–14439.
[19] Stordeur, C., Dalluge, R., Birkenmeier, O., Wienk, H., Rudolph, R., Lange, C. and Lucke, C. (2008) The NMR solution structure of the artificial protein M7 matches the computationally designed model. Proteins 72, 1104–1107.
[20] Hwang, S., Gou, Z. and Kuznetsov, I.B. (2007) DP-Bind: a web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins. Bioinformatics 23, 634–636.
[21] Lim, K., Tempczyk, A., Parsons, J.F., Bonander, N., Toedt, J., Kelman, Z., Howard, A., Eisenstein, E. and Herzberg, O. (2003) Crystal structure of YbaB from *Haemophilus influenzae* (HI0442), a protein of unknown function coexpressed with the recombinational DNA repair protein RecR. Proteins 50, 375–379.
[22] Cooley, A.E., Riley, S.P., Kral, K., Miller, M.C., DeMoll, E., Fried, M.G. and Stevenson, B. (2009) DNA-binding by *Haemophilus influenzae* and *Escherichia coli* YbaB, members of a widely-distributed bacterial protein family. BMC Microbiol. 9, 137.