Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary X-ray crystallographic analysis of a *yedU* gene product from *Escherichia coli*

Ok-Gwan Kim,[a] In-Kwon Kim,[b] Ghyung-Hwa Kim,[a] Junsang Ko,[c] Chankyu Park,[c] Pann-Ghill Suh,[d] Sa-Ouk Kang,[b] Heung-Soo Lee[a]† and Sun-Shin Cha[a]*

[a]Beamline Division, Pohang Accelerator Laboratory, Pohang, Kyungbuk 790-784, South Korea, [b]Laboratory of Biophysics, School of Biological Sciences and Institute of Microbiology, Seoul National University, Seoul 151-742, South Korea, [c]National Creative Research Initiative Center for Behavioral Genetics, Department of Biological Science, Korea Advanced Institute of Science and Technology, Yusong-Ku, Taejon 305-710, South Korea, and [d]Department of Life Science, Division of Molecular and Life Science, Pohang University of Science and Technology, Pohang 790-784, South Korea

† Dr Lee should also be considered as corresponding author. His e-mail address is lhs@postech.ac.kr.

Correspondence e-mail: chajung@postech.ac.kr

Received 16 January 2002
Accepted 22 April 2002

A *yedU* gene product with a molecular mass of 31 kDa is a hypothetical protein with no known function. The protein was purified and crystallized at 296 K. X-ray diffraction data have been collected to 2.3 Å using synchrotron radiation. The crystals belong to the primitive orthorhombic system, with unit-cell parameters $a = 50.56$, $b = 63.45$, $c = 168.02$ Å. The asymmetric unit contains two monomers of the protein, with a corresponding V_M of 2.25 Å^3 Da^{-1} and a solvent content of 44.84%.

1. Introduction

The nucleoid-associated protein H-NS, which is encoded by the gene *hns*, has a central role in the structuring and control of the enteric bacterial chromosome (Williams, 1997). The major role of H-NS is to modulate the expression of a large number of unlinked genes (Atlung & Ingmer, 1997). H-NS regulates the expression of *proVWX*, *bgl*, *appY* and *fimB* of *Escherichia coli* (Atlung *et al.*, 1996; Donato *et al.*, 1997; Higgins *et al.*, 1998) and also affects expression of some virulence genes of *Salmonella typhimurium* and *Shigella* spp. (Maurelli & Sansonetti, 1988; Harrison *et al.*, 1994; O'Byrne & Dorman, 1994). In an *hns*-deletion background, the expression of numerous *E. coli* cellular proteins is either strongly induced or repressed relative to the levels in wild-type cells (Bertin *et al.*, 1990; Yamada *et al.*, 1991). It is thus suspected that H-NS may function as a global transcriptional regulator.

The expression of an *E. coli* 31 kDa protein (*yedU* gene product; YedU) has previously been demonstrated to be greatly enhanced in an *hns*-deletion background (Yoshida *et al.*, 1993). However, no physiological effect of the enhancement or biochemical function of the protein has yet been identified. Although a BLAST (Altschul *et al.*, 1997) database search revealed YedU to be composed of a domain that is evolutionarily conserved from archea to eukarya, the function of the protein still remains obscure. According to PFAM (Bateman *et al.*, 1999), proteins containing the domain are classified as belonging to the ThiJ/PfpI family, which contains many hypothetical proteins. Even the functional members of the family have distinct functions. This family includes proteins involved in the regulation of RNA–protein interactions (Hod *et al.*, 1999), thiamine biosynthesis (ThiJ; EC 2.7.1.49; Mizote *et al.*, 1996), Ras-related signal transduction pathways (Nagakubo *et al.*, 1997), sperm fertility (Welch *et al.*, 1998), *de novo* synthesis of purine nucleotides (FGAM synthase I from *Sulfolobus solfataricus*) and protease activity (Halio *et al.*, 1996) according to the InterPro description of the family (Apweiler *et al.*, 2001). The domain is found at the N-terminus of AraC-like transcriptional regulators (Redenbach *et al.*, 1996) and the C-terminus of HPII-type calatases (Navarro *et al.*, 1996). The domain also appears to be distantly related to glutamine amidotransferases (Du *et al.*, 2000; Horvath & Grishin, 2001). The role of the domain is therefore unclear, which prevents the identification of the biological function of YedU from the primary sequence.

In general, extensive biochemical and physiological studies are required to determine the biological functions of hypothetical proteins. Recently, protein structures have been used to assign the molecular functions of hypothetical proteins, as the function of a gene product is closely related to its three-dimensional structure. Structural homologues of a hypothetical protein are searched for in the Protein Data Bank and the hypothetical protein is then predicted to have molecular functions similar to its homologues. The predictions can provide a basis for assigning a molecular function to the hypothetical protein. For instance, Zarembinski *et al.* (1998), Hwang *et al.* (1999) and Lee *et al.* (2001) have succeeded in structure-based assignment of the molecular function of hypothetical proteins. The overall conformation, detection of the active site and the amino-acid composition of the active site, which can be revealed from the structure of YedU, would lead us to the elucidation of its biochemical function. Here, we report the overexpression, crystallization and

preliminary X-ray crystallographic analysis of YedU as a first step toward its structure determination.

2. Preparation of recombinant YedU

The *yedU* gene coding for a 286 amino-acid polypeptide was amplified by the polymerase chain reaction using *E. coli* genomic DNA as a template. The gene was inserted downstream of the T7 promoter of the expression plasmid pET-21b (Novagen) and the plasmid was introduced into *E. coli* strain BL21 (DE3). Cells were grown to an OD_{600} of approximately 0.5 in Luria–Bertani media containing 0.1 mg ml^{-1} ampicillin (Duchefa) at 310 K and the expression of YedU was induced by 1 m*M* isopropyl-β-D-thiogalactoside (Duchefa). After 3 h induction, cells were harvested and resuspended in 20 m*M* sodium phosphate pH 7.6 containing 1 m*M* dithiothreitol (Sigma). After sonication and subsequent centrifugation, the supernatant was loaded onto a nickel–nitrilotriacetic acid column (Qiagen) and the protein was eluted with a 0–500 m*M* imidazole (Fluka) gradient in 20 m*M* sodium phosphate pH 7.6 (Fluka). YedU was further purified using a Superdex 200 HR 16/60 molecular-sizing column (Amersham Pharmacia Biotech) with 20 m*M* sodium phosphate pH 7.6 and was then concentrated to approximately 12 mg ml^{-1} for crystallization.

3. Crystallization and X-ray analysis

Crystals of YedU were obtained by the hanging-drop vapour-diffusion method using 24-well Costa plates at 296 K. The first crystallization screening was performed with Crystal Screen, a sparse-matrix screening kit (Hampton Research, USA). A number of microcrystals were produced in a condition containing 30% polyethyleneglycol 1500 (PEG 1.5K, Fluka) in 1 d. Subsequently, the initial crystallization condition was optimized to produce larger single crystals in droplets containing 2 µl of protein sample (12 mg ml^{-1}) and an equal volume of precipitant solution containing 26% PEG 1.5K. The droplets were equilibrated against 1 ml of the same precipitant solution at 296 K and the crystals grew to maximum size in a week (0.01 × 0.01 × 0.1 mm; Fig. 1). For data collection, crystals were frozen at 100 K using a Cryostream cooler (Oxford Cryosystems) after briefly being immersed in a cryoprotectant solution containing 15% glycerol in the same precipitant solution. A 2.3 Å data set (Table 1) was obtained using a MacScience DIP2030b imaging plate at beamline 6B at the Pohang Light Source (PLS), Korea. Using an autoindexing program provided with the program *HKL* (Otwinowski, 1993) and examining the diffraction data set, we found that the crystals belong to the primitive orthorhombic system. Two molecules of YedU were contained in the asymmetric unit, corresponding to a crystal volume per unit molecular weight (V_M) of 2.25 Å^3 Da^{-1}, with a solvent content of 44.84% (Matthews, 1968). However, NCS (non-crystallographic symmetry) was not detectable from a self-rotation function. This happens, for instance, if the NCS runs parallel to a crystallographic symmetry axis or if the NCS is purely translational.

Table 1
Crystal information and data-collection statistics.

Values in parentheses refer to the highest resolution shell, 2.30–2.38 Å.

Source	6B, PLS
Wavelength (Å)	1.009
Space group	*P*222 or $P2_12_12$
Unit-cell parameters (Å)	a = 50.6, b = 63.5, c = 168
Resolution range (Å)	10.0–2.3
Completeness (>1σ, %)	95.4 (78.6)
R_{sym}† (%)	9.4 (33.7)
$I/\sigma(I)$	15.26 (2.8)

† $R_{sym} = \sum |I_{obs} - I_{avg}|/I_{obs}$.

4. Discussion

Two members of the ThiJ/PfpI family, *Pyrococcus horikoshii* intracellular protease PH1704, with 90% sequence identity to *P. furiosus* PfpI (referred to hereafter as PfpI), and *E. coli* HPII catalase, have been studied structurally (Bravo *et al.*, 1995; Du *et al.*, 2000). PfpI and the C-terminal ThiJ/PfpI domain of the catalase show 25.9% (Fig. 2) and 19.5% sequence identity to YedU, respectively. Interestingly, two catalytic residues (Cys100 and His101) of the protease are well conserved in YedU, while the C-terminal ThiJ/PfpI domain of the catalase lacks these functional residues. The sequence around the cysteine residue in both proteins (SLCHGP in YedU and SICHGP in the protease) is consistent with the consensus sequence small-*x*-Nu-*x*-small-small for α/β hydrolases (Du *et al.*, 2000), indicating that the two proteins may be closely related in structure and function. The crystal structure of *P. horikoshii* PH1704 reveals that PfpI-type proteases are homohexameric proteases, which is consistent with the experimental observation that activity was found only for the oligomeric forms of the protein (Du *et al.*, 2000). Molecular replacement was attempted using *AMoRe* (Navaza, 1994) and *CNS* (Brünger *et al.*, 1998) with monomeric and dimeric models generated from the hexameric PfpI

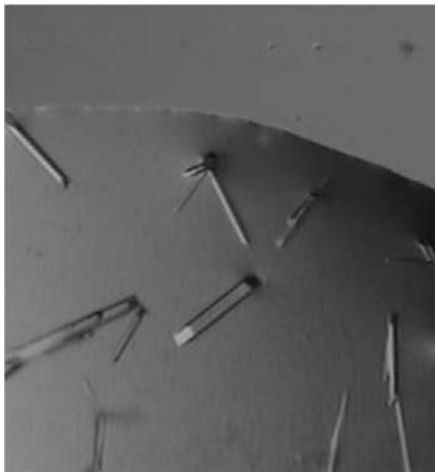

Figure 1
Orthorhombic crystals of YedU from *E. coli*.

```
YedU       MTVQTSKNPQVDIAEDNAFFPSEYSLSQYTSPVSDLDGVDYPKPYRGKHKILVIAADERY
PfpI       ------------------------------------------------------------

YedU       LPTDNGKLFSTGNHPIETLLPLYHLHAAGFEFEVATISGLMTKFEYWAMPQKDEKVMPFF
PfpI       ----MKVLFLTAN-EFEDVELIYPYHRLKEEGHEVYIA---S-FERGTITGK--------
                 ** *.*  :* :  :*  *    * . . *:   : **  ::. *

YedU       EQHKSLFRNPKKLADVVASLNADSEYAAIFVPGGHGALIGLPESQDVAAALQWAIKNDRF
PfpI       --HGYSV----KVDLTFDKVNPE-EFDALVLPGGR-APERVRLNEKAVSIARKMFSEGKP
             *   .    *:  .. .:*.: *: *:.:***: *   :  .:...:  :  :.:.:

YedU       VISLCHGPAAFLALRHGDNPLNGYSICAFPDAADKQTPEIGYMPGHLTWYFGEELKKMGM
PfpI       VASICHGPQILISA--G--VLRGRKGTSYPGIKDDMINAG------VEWVD-AEVVVDG-
           * *:****  :::   *   *.* .  ::*.  *.          : *    *:   *

YedU       NIINDDITGRVHKDRKLLTGDSPFAANALGKLAAQEMLAAYAG
PfpI       N----WVSSRVPAD--LYAWMREFVK--LLK-----------
           *     ::.**  *  * :    *.   * *
```

Figure 2
CLUSTALW sequence alignment between YedU and PfpI. PfpI exhibits the highest sequence similarity (25.9% identity) to YedU among ThiJ/PfpI family members. *, : and . indicate perfect match, high similarity and low similarity, respectively.

structure. However, all trials resulted in failure. The hexameric conformation of the protease (a dimer of trimers) cannot be applied to YedU because of the orthorhombic symmetry of the YedU crystal with two monomers in an asymmetric unit. The formation of the catalytic triad in PfpI is based on the hexameric conformation: Cys100 and His101 form the triad with a glutamate residue from an adjacent monomer. As YedU appears to have a distinct oligomeric conformation, it cannot form the same catalytic triad as PfpI. Therefore, the quaternary structures and local conformation of the active sites of YedU and PfpI must be quite different from each other. In addition, YedU has about 100 more amino acids than PfpI, most of which are at the N-terminus of YedU (Fig. 2). Therefore, we cannot precisely infer the structure and biochemical function of YedU from those of PfpI. As an approach to assigning the molecular structure and function of YedU, efforts toward structure determination using MAD with SeMet-YedU crystals are in progress. Structural comparison between YedU and other members of the ThiJ/PfpI family will also contribute to determining the evolutionary relationship between ThiJ/PfpI family members of diverse biological functions that share the common domain.

This study was supported by a grant from the Korea Health 21 R&D Project, Ministry of Health & Welfare (HMP-00-CH-13-0012) and in part by grants from The Center for Biological Modulators and National Creative Research Initiative Center for Behavioral Genetics.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.

Apweiler, R. *et al.* (2001). *Nucleic Acids Res.* **29**, 37–40.

Atlung, T. & Ingmer, H. (1997). *Mol. Microbiol.* **24**, 7–17.

Atlung, T., Sund, S., Olesen, K. & Brondsted, L. (1996). *J. Bacteriol.* **178**, 3418–3425.

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Finn, R. D. & Sonnhammer, E. L. L. (1999). *Nucleic Acids Res.* **27**, 260–262.

Bertin, P., Lejenue, P., Laurent-Winter, C. & Danchin, A. (1990). *Biochimie*, **72**, 889–891.

Bravo, J., Verdaguer, N., Tormo, J., Betzel, C., Switala, J., Loewen, P. C. & Fita, I. (1995). *Structure*, **15**, 491–502.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. M., Gros, P., Grosse-Kunstleve, R .W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Donato, G. M., Lilivelt, M. J. & Kawula, T. H. (1997). *J. Bacteriol.* **179**, 6618–6625.

Du, X., Choi, I. G., Kim, R., Wang, W., Jancarik, J., Yokota, H. & Kim, S.-H. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 14079–14084.

Halio, S. B., Blumentals, I., Short, S. A., Merrill, B. M. & Kelly, R. M. (1996). *J. Bacteriol.* **178**, 2605–2612.

Harrison, J. A., Pickard, D., Higgins, C. F., Khan, A., Chatfield, S. N., Ali, T., Dorman, C. J., Hormaeche, C. E. & Dougan, G. (1994). *Mol. Microbiol.* **13**, 133–140.

Higgins, C. F., Dorman, C. J., Stirling, D. A., Waddell, L., Booth, I. R., May, G. & Bremer, E. (1998). *Cell*, **52**, 559–584.

Hod, Y., Pentyala, S. N., Whyard, T. C. & El-Maghrabi, M. R. (1999). *J. Cell Biochem.* **72**, 435–444.

Horvath, M. M. & Grishin, N. V. (2001). *Proteins Struct. Funct. Genet.* **42**, 230–236.

Hwang, K. Y., Chung, J. H., Kim, S. H., Han, Y. S. & Cho, Y. (1999). *Nature Struct. Biol.* **6**, 422–426.

Lee, J. Y., Kwak, J. E., Moon, J., Eom, S. H., Liong, E. C., Pedelacq, J.-D., Berendzen, J. & Suh, S. W. (2001). *Nature Struct. Biol.* **8**, 789–794.

Matthews, B. (1968). *J. Mol. Biol.* **33**, 491–497.

Maurelli, A. T. & Sansonetti, P. J. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 564–576.

Mizote, T., Tsuda, M., Nakazawa, T. & Nakayama, H. (1996). *Microbiology*, **142**, 2969–2974.

Nagakubo, D., Taita, T., Kitaura, H., Ikeda, M., Tamai, K., Iguchi-Ariga, S. M. & Ariga, H. (1997). *Biochem. Biophys. Res. Commun.* **13**, 509–513.

Navarro, R. E., Stringer, M. A., Hansberg, W., Timberlake, W. E. & Aguirre, J. (1996). *Curr. Genet.* **29**, 352–359.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

O'Byrne, C. P. & Dorman, C. J. (1994). *FEMS Microbiol. Lett.* **121**, 99–105.

Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

Redenbach, M., Kieser, H. M., Denapaite, D., Eichner, A., Cullum, J., Kinashi, H. & Hopwood, D. A. (1996). *Mol. Microbiol.* **21**, 77–96.

Welch, J. E., Barbee, R. R., Roberts, N. L., Suarez, J. D. & Klinefelter, G. R. (1998). *J. Androl.* **19**, 385–393.

Williams, R. M. (1997). *FEMS Microbiol Lett.* **156**, 175–185.

Yamada, H., Yoshida, T., Tanaka, K., Sanakawa, C. & Mizuno, T. (1991). *Mol. Gen. Genet.* **230**, 332–336.

Yoshida, T., Ueguchi, C., Yamada, H. & Mizuno, T. (1993). *Mol. Gen. Genet.* **237**, 113–122.

Zarembinski, T. I., Hung, L. W., Mueller-Dieckmann, H. J., Kim, K. K., Yokota, H., Kim, R. & Kim, S. H. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 15189–15193.