

Jeyaraman Jeyakanthan, Eiji
Inagaki, Chizu Kuroishi and
Tahir H. Tahirov*Advanced Protein Crystallography Research
Group, RIKEN Harima Institute, 1-1-1 Kouto,
Mikazuki-cho, Sayo-gun, Hyogo 679-5148,
Japan

Correspondence e-mail: tahir@spring8.or.jp

Received 4 March 2005

Accepted 20 April 2005

Online 26 April 2005

PDB References: PH0500-I, 1v96, r1v96sf;
PH0500-II, 1ye5, r1ye5sf.

Structure of PIN-domain protein PH0500 from *Pyrococcus horikoshii*

The *Pyrococcus horikoshii* OT3 protein PH0500 is highly conserved within the *Pyrococcus* genus of hyperthermophilic archaea and shows low amino-acid sequence similarity with a family of PIN-domain proteins. The protein has been expressed, purified and crystallized in two crystal forms: PH0500-I and PH0500-II. The structure was determined at 2.0 Å by the multiple anomalous dispersion method using a selenomethionyl derivative of crystal form PH0500-I (PH0500-I-Se). The structure of PH0500-I has been refined at 1.75 Å resolution to an *R* factor of 20.9% and the structure of PH0500-II has been refined at 2.0 Å resolution to an *R* factor of 23.4%. In both crystal forms as well as in solution the molecule appears to be a dimer. Searches of the databases for protein-fold similarities confirmed that the PH0500 protein is a PIN-domain protein with possible exonuclease activity and involvement in DNA or RNA editing.

1. Introduction

Genomic projects have revealed the whole genome sequences of various organisms. However, the resulting analysis indicates that no known function can be inferred for a significant fraction of the gene products. To infer functions, additional information beyond sequences is needed. Since the biological function of a gene product is tightly coupled with its three-dimensional structure, finding the structure or its folding pattern may provide an important insight into the function of the gene product. The archaea are an attractive target for structural genomics projects because of evolutionary aspects, their small genome size and their thermophilic nature. Whole genome sequences have been determined from some species of archaea. The whole genome sequence of *Pyrococcus horikoshii* OT3 reveals 2061 open reading frames (ORFs), many of which would express proteins of unknown function (Kawarabayashi *et al.*, 1998). The amino-acid sequence of PH0500 from *P. horikoshii* OT3, a target of current structural studies, is well conserved in the *Pyrococcus* genus, including *P. furiosus* (Cohen *et al.*, 2003) and *P. abyssi* (Robb *et al.*, 2001) (Fig. 1), and exhibits low homology to PIN domains, named after their homology with the N-terminal domain of the pili biogenesis protein (PIN, PilT N-terminus; Wall & Kaiser, 1999), comprise a large family of proteins with over 300 members

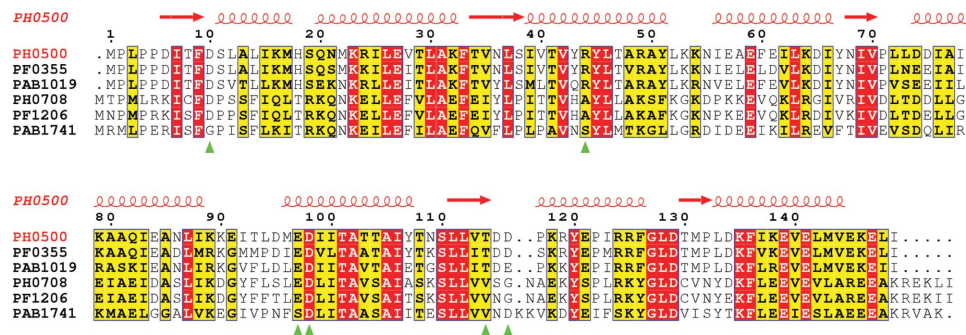
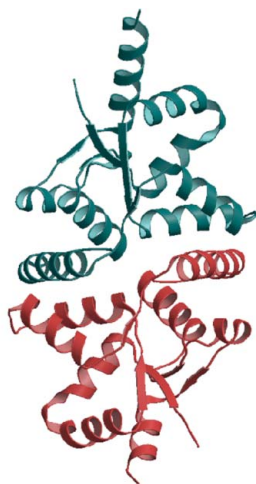


Figure 1

Amino-acid sequence alignment of PH0500 with five hypothetical proteins from hyperthermophilic archaea. Identical residues are highlighted in red. Secondary-structure elements are shown above the alignment. The active-site residues are highlighted with green triangles according to PH0500 sequence numbering. These proteins are from *P. horikoshii* (PH0500, PH0708), *P. furiosus* (PF0355, PF1206) and *P. abyssi* (PAB1019, PAB1741). Alignment was performed using CLUSTALW (Thompson *et al.*, 1994) and the figure was produced using ESPript (Gouet *et al.*, 1999).

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the last shell.

	PH0500-I	PH0500-II	PH0500-I-Se
Wavelength (Å)	1.0000	1.0000	0.9792, 0.9795, 1.0000
Space group	$P2_12_12_1$	$P2_1$	$P2_12_12_1$
Unit-cell parameters			
<i>a</i> (Å)	44.91	54.34	45.81
<i>b</i> (Å)	95.46	46.38	96.63
<i>c</i> (Å)	102.66	62.49	104.88
β (°)	90	113.33	90
V_M (Å ³ Da ⁻¹)	3.2	2.1	3.4
<i>Z</i>	2	2	2
Solvent content (%)	61.4	40.9	63.5
Resolution (Å)	20.0–1.75 (1.81–1.75)	20.0–2.0 (2.07–2.0)	20.0–2.0 (2.07–2.0)
No. of observations	142329	64492	133520, 133562, 133752
No. of unique reflections	45126	18805	30453, 30450, 30472
Completeness (%)	99.2 (99.9)	95.6 (95.9)	99.3 (100), 99.3 (100), 99.4 (100)
R_{merge} (%)	3.9 (19.9)	6.0 (21.0)	6.4 (22.6), 4.7 (21), 3.8 (22)
$\langle I/\sigma(I) \rangle$ (%)	16.3(5.2)	13.5 (5.9)	16.2 (6.8), 15.7 (6.3), 15.3 (5.9)
Figure of merit (<i>SOLVE/RESOLVE</i>)			0.36/0.57
Protein atoms	2349	2337	
Solvent atoms	384	217	
<i>R</i>	20.9	23.4	
R_{free}	22.8	27.3	
R.m.s. deviations from ideal values†			
Bond lengths (Å)	0.01	0.01	
Bond angles (°)	2.50	1.2	
Dihedral angles (°)	20.20	19.9	
Impropers (°)	0.88	0.77	
Residues (%) in Ramachandran plot‡			
Core region	94.4	94.1	
Additionally allowed region	5.6	5.5	
Generously allowed region	0.0	0.4	
Disallowed region	0.0	0.0	

† Deviations from ideal geometry parameters as defined by Engh & Huber (1991). ‡ As calculated by *PROCHECK* (Laskowski *et al.*, 1993).

found in eukaryotes, bacteria and archaea. PIN domains were initially thought to function in signalling (Noguchi *et al.*, 1996), but recent bioinformatics analysis has suggested the presence of a set of five highly conserved acidic residues and a sixth conserved position where there is either a serine or threonine residue. Such amino-acid conservation occurs in enzymes, in particular those, such as nucleases, that ligate divalent cations. This analysis has led to a suggested exonuclease function (Clissold & Ponting, 2000).

Current structural studies confirm that PH0500 belongs to a wider family of PIN-domain-containing proteins (Clissold & Ponting, 2000) and in spite of low amino-acid sequence homology, exhibits structural similarities to PIN-domain proteins from *Pyrobaculum aerophilum* (PAE2754; Arcus *et al.*, 2004) and *Archaeoglobus fulgidus* (AF0591; Levin *et al.*, 2004). The protein PAE2754 has been shown to exhibit exonuclease activity with possible involvement in DNA repair (Arcus *et al.*, 2004). Here, we report the crystal structure of PH0500 and compare it with the structures of PAE2754 and AF0591, highlighting the similarities and unique features of these archaeal PIN-domain proteins.

2. Material and methods

2.1. Expression and purification

The gene was amplified by the polymerase chain reaction (PCR) using *P. horikoshii* OT3 genomic DNA as a template. The recombinant plasmid was constructed by the super-rare-cutter system (Kanagawa *et al.*, manuscript in preparation). *Escherichia coli* BL21-CodonPlus (DE3)-RIL cells were transformed with the recombinant plasmid and grown at 310 K in LB medium containing 50 µg ml⁻¹ ampicillin for 20 h. The cells were harvested by centrifugation at 6500 rev min⁻¹ for 5 min, suspended in 20 mM Tris-HCl pH 8.0

(buffer A) containing 0.5 M NaCl and 5 mM 2-mercaptoethanol and disrupted by sonication. The cell lysate was heated at 363 K for 13 min. After heat treatment, denatured proteins were removed by centrifugation (15 000 rev min⁻¹, 30 min) and the supernatant solution was used as the crude extract for purification. The crude extract was desalted using a HiPrep 26/10 desalting column (Amersham-Biosciences) and applied onto a Super Q Toyopearl 650M column (Tosoh) equilibrated with buffer A. The protein was eluted with a linear gradient of 0–0.3 M NaCl in buffer A. The protein was desalted with a HiPrep 26/10 desalting column with buffer A and loaded onto a Resource Q column (Amersham Biosciences) equilibrated with buffer A. The protein was eluted with a linear gradient of 0–0.3 M NaCl in buffer A. The buffer of the fractions containing the protein was exchanged to 10 mM sodium phosphate pH 7.0 using a HiPrep 26/10 desalting column and applied onto a Bio-Scale CHT-20-I column (Bio-Rad) equilibrated with the same buffer. The protein was eluted with a linear gradient of 10–200 mM sodium phosphate pH 7.0. The fractions containing protein were pooled, concentrated by ultrafiltration (Vivaspin, 5 kDa cutoff) and loaded onto a HiLoad 16/60 Superdex 75pg column (Amersham Biosciences) equilibrated with buffer A containing 0.2 M NaCl. The purified protein was homogeneous on native PAGE.

2.2. Dynamic light-scattering measurements

A dynamic light-scattering experiment was performed on a DynaPro MS/X instrument from Protein Solutions (Lakewood, New Jersey). The measurements were made at 291 K on the purified protein at 0.5–1 mg ml⁻¹ in solution containing 20 mM Tris-HCl buffer pH 8.0 and 200 mM sodium chloride.

2.3. Crystallization

The crystallization was performed by the sitting-drop vapour-diffusion technique at 293 K using Linbro multiwell plates. Each drop, consisting of 1 μl of a 20 mg ml⁻¹ protein solution in 0.1 M Tris-HCl buffer pH 7.9 and 1 μl reservoir solution, was allowed to equilibrate against 500 μl reservoir solution. Preliminary crystallization conditions were established using Hampton Research Crystal Screens I and II (Jancarik & Kim, 1991), followed by a refinement of the conditions through variation of protein concentration, precipitant, pH, drop volume and additives. The best conditions yielded crystals of two different forms. Crystals of form I (PH0500-I) were grown using a reservoir solution consisting of 6% (w/v) PEG 4000, 0.1 M sodium citrate buffer pH 5.2 and 25% (v/v) glycerol. The crystal appeared within 1 d and reached final dimensions of 0.3 \times 0.2 \times 0.1 mm after 3–5 d. They diffracted X-rays to a resolution of 1.75 \AA . Crystals of form II (PH0500-II) were produced with reservoir solution containing 12.5% (w/v) PEG 4000, 0.1 M MES buffer pH 5.8 and 30% (v/v) glycerol. The crystals grew to approximate dimensions of 0.2 \times 0.1 \times 0.1 mm within 3 d and diffracted to a resolution of 2.0 \AA . The Matthews coefficient (V_M) values of each crystal were estimated to be 3.2 $\text{\AA}^3 \text{Da}^{-1}$ for PH0500-I and 2.1 $\text{\AA}^3 \text{Da}^{-1}$ for PH0500-II, respectively, assuming that two molecules are contained in the asymmetric units (Matthews, 1968). The selenomethionyl derivative of PH0500-I crystal (PH0500-I-Se) was obtained under the same conditions as used for the native PH0500-I crystals and diffracted X-rays to a resolution of 2.0 \AA . Crystallographic parameters are summarized in Table 1.

2.4. Data collection and processing

Diffraction data were collected using a Rigaku R-AXIS V imaging-plate detector at the BL26B1 beamline, SPring-8, Japan. The crystals were flash-frozen in a nitrogen-gas stream at 90 K for data collection with oscillation angles of 1.0° and a crystal-to-detector distance of 250 mm. The X-ray absorption spectrum at the Se *K* edge was measured from the PH0500-I-Se crystal. The same crystal was used for multiwavelength anomalous dispersion (MAD) data collection. Based on the fluorescence spectrum, two energy levels were chosen for the data collection, both of which were near the absorption edge of the Se atom: 1266.18 keV ($\lambda = 0.97920 \text{ \AA}$) and 1265.75 keV ($\lambda = 0.97950 \text{ \AA}$). The third energy level was set to 1239.84 keV ($\lambda = 1.0000 \text{ \AA}$) as a remote point. The wavelength was set to 1.0000 \AA for the diffraction data collection from native crystals PH0500-I and PH0500-II. The diffraction data were processed and scaled with the

HKL2000 package (Otwinowski & Minor, 1997). The diffraction data-collection statistics are summarized in Table 1.

2.5. Structure determination and refinement

The structure was solved using the MAD method (Hendrickson *et al.*, 1990). All selenium sites, except for the N-terminal site, were found and refined and the initial phases were calculated with the programs *SOLVE* and *RESOLVE* (Terwilliger & Berendzen, 1999). The phases were improved by solvent flattening and histogram matching and used to automatically build a partial model of the protein with the program *ARP/wARP* (Perrakis *et al.*, 1999). Unambiguous parts and the side chains were added during the refinement, without non-crystallographic symmetry (NCS) restraints. The rest of the residues were built manually using *QUANTA* (Accelrys). All crystallographic refinements were carried out using *CNS* v1.1 (Brünger *et al.*, 1987, 1998). Solvent molecules were gradually included into the structure at stereochemically sensible positions and with difference density higher than 3.0σ and $2F_o - F_c$ density higher than 0.8σ . The PH0500-I-Se structure was used as a starting model for the refinement of PH0500-I structure at 1.75 \AA . The crystal structure of PH0500-II was solved by the molecular-replacement method with the program *AMoRe* (Navaza, 1994) using the PH0500-I structure as a search model. The PH0500-II structure was refined at 2.0 \AA resolution. A summary of structure-determination and refinement statistics is given in Table 1.

3. Results and discussion

3.1. Quality of the structures

The structure of the PH0500-I-Se crystal was determined by the MAD method at 2.0 \AA resolution. The native PH0500-I structure was refined at 1.75 \AA resolution to an *R* factor and R_{free} of 20.9 and 22.8%, respectively. The final model for PH0500-I includes 293 amino acids (residues 2–49 and 52–149 for molecule *A* and residues 2–145 for molecule *B*), 366 water molecules and three glycerol molecules. Residues 50–58 are disordered, with a complete absence of electron density for residues 50 and 51 in subunit *A*, whereas in subunit *B* the residues in those regions are ordered with well defined electron density. The electron density is well ordered in both subunits of PH0500-II when compared with that of the PH0500-I protomers. The PH0500-II structure was solved by the molecular-replacement method and refined at 2.0 \AA resolution to an *R* factor and R_{free} of 23.4 and 27.3%, respectively, and contains 289 amino acids (residues 2–143 for subunit *A* and residues 2–149 for subunit *B*) and 217 water molecules. In both native structures, PH0500-I and PH0500-II, the main-chain φ and ψ angles of over 94% of residues lie within the most favoured region in the Ramachandran plot, indicating the excellent geometry of the refined structures (Laskowski *et al.*, 1993). Superposition of C α atoms of the monomer subunits of PH0500 with PAE2754 and AF0591 monomers gave root-mean-square deviations of 2.4 and 2.9 \AA , respectively, and is shown in Fig. 5.

3.2. Overall structure

Both crystal structures, PH0500-I and PH0500-II, revealed one homodimer with a non-crystallographic twofold symmetry consistent with

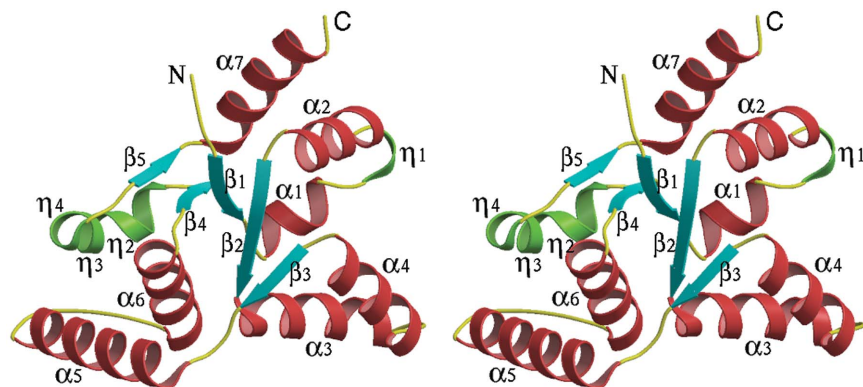


Figure 2
Stereo representation of PH0500 subunit ribbon diagram with labelling of secondary-structure elements. The figure was prepared using *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt & Murphy, 1994).

solvent-content calculations (Table 1). The dynamic light-scattering experiments confirmed the presence of dimers in solution. Each subunit consists of a central twisted β -sheet of five short parallel strands with order $\beta 3$, $\beta 2$, $\beta 1$, $\beta 4$ and $\beta 5$ flanked with seven α -helices and four one-turn 3_{10} -helices, $\eta 1$ – $\eta 4$ (Fig. 2). Helices $\alpha 1$ – $\alpha 4$, $\alpha 7$ and $\eta 1$ are located on one side of the β -sheet and helices $\alpha 5$, $\alpha 6$ and $\eta 3$ – $\eta 5$ on the opposite side of the β -sheet. Helix $\eta 1$ forms part of loop $\alpha 1\alpha 2$, while the three adjacent and successive helices $\eta 2$ – $\eta 4$ form an arc-like helical element connecting strands $\beta 4$ and $\beta 5$. The hydrophobic cores on either side of the β -sheet are extended toward the homodimerization interface and contribute to the formation of a homodimer with a large buried surface area of over 2400 Å² (Fig. 3a). Such a large buried surface area is characteristic of biologically functional stable dimers (Janin, 1997). The intersubunit interface is formed by helices $\alpha 3$, $\eta 2$, $\alpha 4$ and $\alpha 5$, the N-terminus of helix $\alpha 6$ and the loop $\beta 3\alpha 5$ of each subunit. 15 residues (Ile39, Val40, Tyr43, Tyr51, Ala50, Ile56, Phe60, Leu72, Ala77, Ala80, Ala81, Ala85, Ile88, Met96 and Ile100) from each subunit are involved in formation of the intramolecular hydrophobic core; however, few intersubunit hydrogen bonds, including ion pairs and the bridging water molecules, contribute to dimer formation. The sequence alignment of PH0500 with homologous proteins from the *Pyrococcus* genus (Fig. 1) reveals that the surfaces of these proteins corresponding to the homodimerization surface of PH0500 are also mainly hydrophobic. This indicates that the *Pyrococcus* proteins listed in Fig. 1 may form dimeric structures similar to that of PH0500. The PH0500 molecule is rhomb-shaped and has an extended charged depression on one side. Both the depression-containing side and the opposite side of the molecule are

negatively charged on the central areas and positively charged on the surrounding areas (Fig. 4).

3.3. Database searches and classification of the fold

The structure determination of PH0500 by the MAD method enabled the use of various search tools in order to classify the fold of this protein. The polypeptide fold of PH0500 was used to find matches in the PDB by executing searches with DALI (Holm & Sander, 1993) and CATH (Pearl *et al.*, 2000) and the topology was also used for a TOPS domain-database search (Michalopoulos *et al.*, 2004). The best structural match with a Z score of 12.0 was found with a DALI search for a PIN domain of PAE2754 from *Pyrobaculum aerophilum* (Arcus *et al.*, 2004), which shares only 12% sequence identity with PH0500, although the topology of the PH0500 fold appears to be very similar to the topology of the PAE2754 PIN-domain fold. Second and third on the list of the DALI search were domain B of tryptophanyl tRNA synthetase (Ilyin *et al.*, 2000) and biotin carboxylase protein (Waldrop *et al.*, 1994) with Z scores of 4.7 and 4.6, respectively. However, the database searches described above failed to identify the structure of a recently reported PIN-domain protein from *Archaeoglobus fulgidus*, AF0591 (Levin *et al.*, 2004), probably owing to the differences in secondary structure described below.

3.4. Comparison with PAE2754 and AF0591 structures

Comparison of PH0500 with PAE2754 and AF0591 revealed several significant differences in folding, size and positioning of

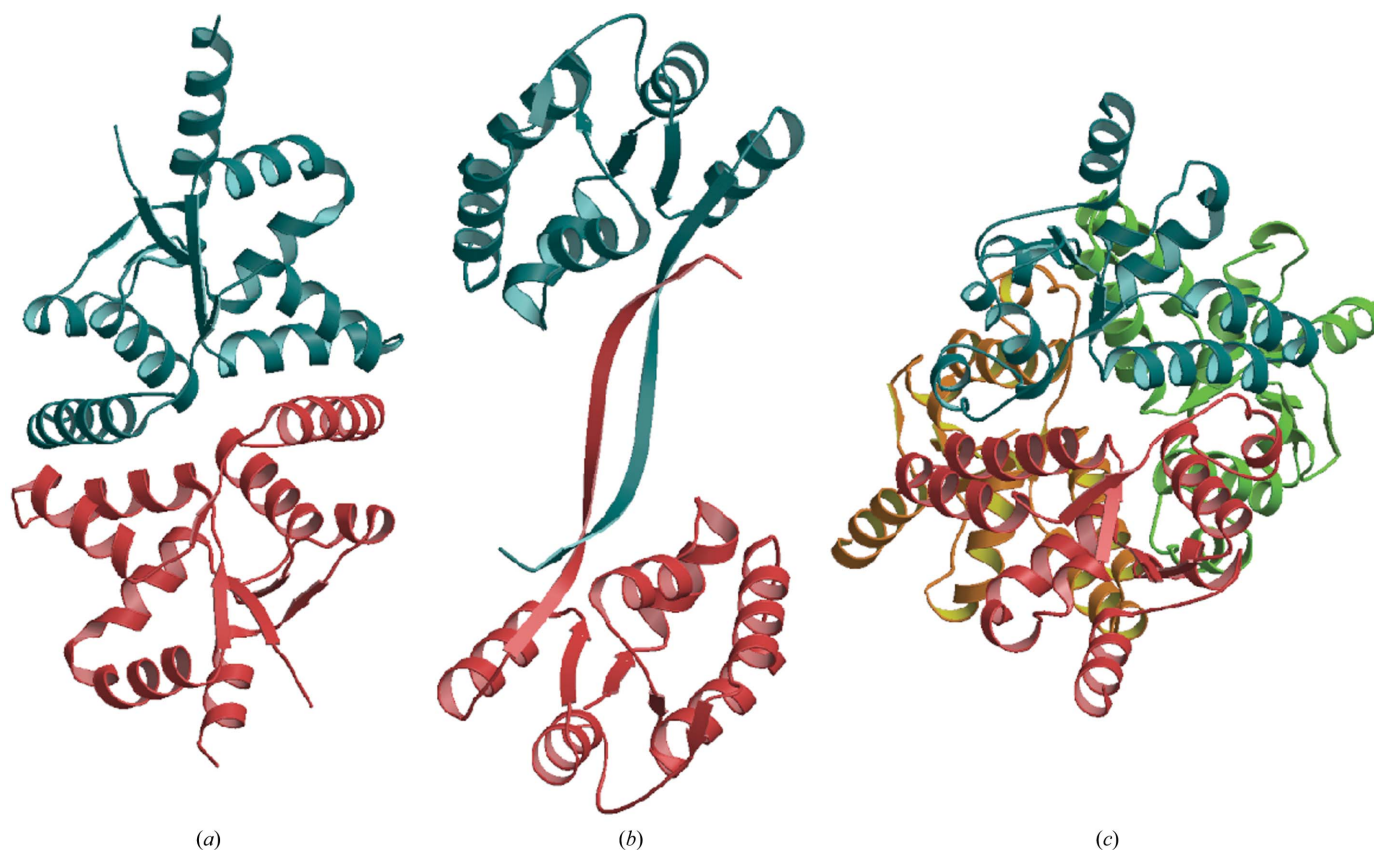


Figure 3 The overall view of (a) PH0500 and (b) AF0591 dimers and (c) PAE2754 tetramers. Subunits are represented by ribbons and have different colours. The figures were prepared using *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt & Murphy, 1994).

secondary-structure elements (Figs. 3 and 5). In particular, the 3_{10} -helix η_1 was absent and the 3_{10} -helices η_2 – η_4 were replaced by a short α -helix in both PAE2754 and AF0591. However, the most dramatic differences were observed for the elements contributing to dimerization. For example, helix α_5 of PH0500 was absent in AF0591 and the axis of PH0500 helix α_4 was rotated nearly 55 and 30° relative to the corresponding helices in AF0591 and PAE2754 structures, respectively, although in the AF0591 structure the sequences corresponding to helix α_7 in PH0500 were folded as an extension of strand β_5 and participated in formation of dimers in the crystal. The significance of AF0591 dimerization has not yet been reported; however, the involvement of the same sequences in the formation of either α -helix or β -strand may depend on interacting partners. For example, such a ‘chameleon’ fragment has been

discovered in the structure of the MAT α 2–MCM1–DNA complex (Tan & Richmond, 1998).

Because of the above-described differences in the elements involved in dimer formation, the overall structures of PH0500, PAE2754 and AF0591 dimers differ significantly, resulting in different shapes of the depressions containing the putative active sites as well as in different relative orientations and separations of the active sites by 28, 25 and 42 Å, respectively (Fig. 3). Moreover, the dimers of PAE2754 dimerize to form a stable tetramer both in the crystal and in solution. Such tetramer formation brings all four active sites into a negatively charged channel between the dimers (Arcus *et al.*, 2004).

The structure of PAE2754 has four conserved acidic residues in the putative active site in each monomer, Asp8, Glu38, Asp92 and Asp110, which are clustered together in a surface pocket, facing into the tunnel through the centre, and a conserved Thr108, which is hydrogen bonded to Asp8. The corresponding residues in AF0591 are Asp17, Glu52, Asp88, Asp106 and Thr104. Each monomer structure of PH0500 has three out of the four conserved active-site acidic residues (Asp10, Asp98 and Asp116) as well as a conserved threonine (Thr114), whereas the Glu38 residue of PAE2754 is replaced by Arg44 in PH0500. However, the function of PAE2754 Glu38 in PH0500 seems to be rescued by Glu97, the carboxyl group of which occupies a position similar to the carboxyl group position of Glu38 in the PAE2754 structure (Fig. 5). The conservation of three-dimensional positions of the four acidic residues and threonine in the putative active sites of the PIN-domain proteins PAE2754, AF0591 and PH0500 (Fig. 5) indicates that these three proteins may have similar catalytic properties.

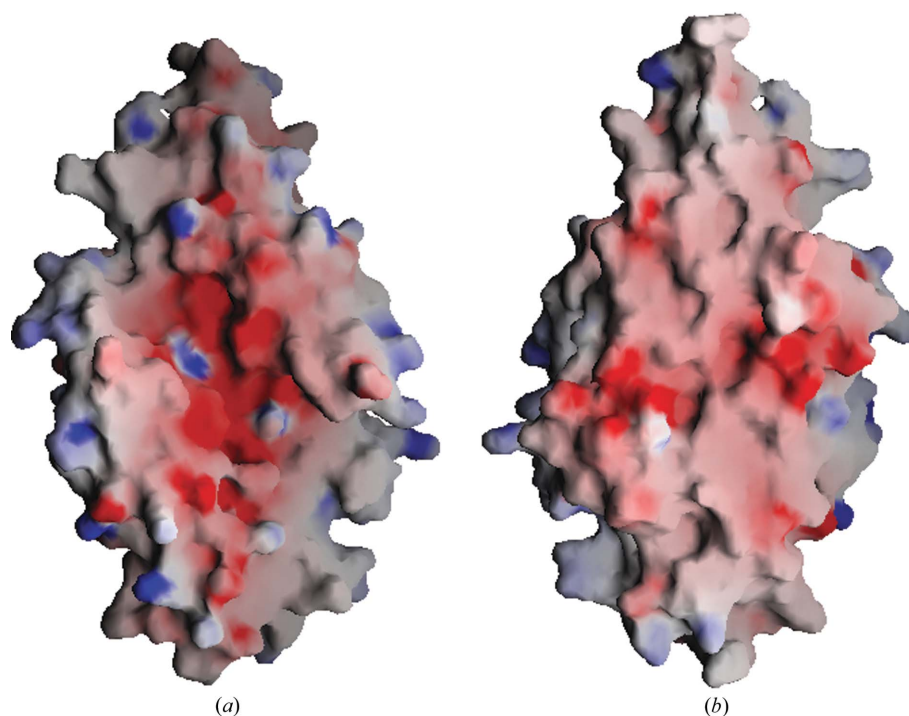


Figure 4

The surface representation of the PH0500 dimer. View (a) is as in Fig. 3(a) and the view (b) is in the opposite direction to (a). The positively and the negatively charged surface regions are in blue and red, respectively. The figures were prepared using GRASP (Nichols *et al.*, 1991).

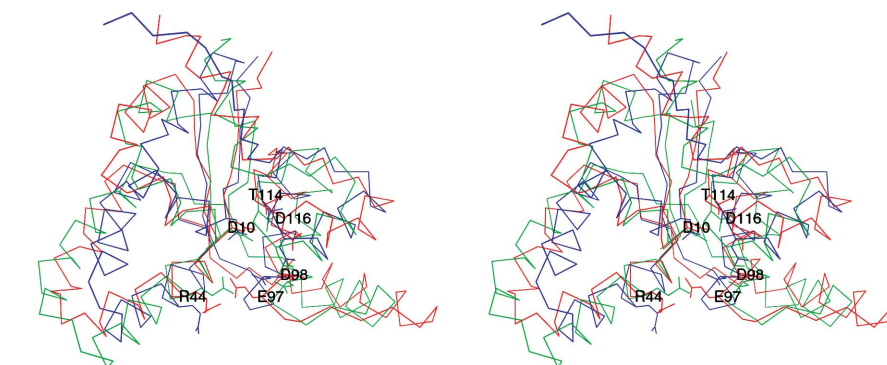


Figure 5

Comparison of the subunit structures of PH0500, AF0591 and PAE2754. Stereoview of superimposed C α traces of PH0500, AF0591 and PAE2754 are represented by red, blue and green lines, respectively. The side chains of the conserved active-site residues are also shown and labelled according to PH0500 sequence numbering. The figure was prepared using MOLSCRIPT (Kraulis, 1991).

4. Conclusions

The PIN domains (Wall & Kaiser, 1999) comprise a large family of proteins with over 300 members found in eukaryotes, bacteria and archaea. Recent bioinformatic analysis of PIN domains suggested their catalytic activity as a magnesium-dependent 5'-exonuclease with a possible involvement in nonsense-mediated mRNA decay and RNAi in eukaryotes (Clissold & Ponting, 2000). Indeed, the exonuclease activity has been confirmed for the archaeal PIN domain PAE2754, suggesting its involvement in DNA repair (Arcus *et al.*, 2004). The conservation of the key active-site residue positions in AF0591 and PH0500 indicates that the latter two proteins may also possess exonuclease activity. However, the above-described significant differences in oligomerization and overall shapes of PAE2754, AF0591 and PH0500 molecules suggest that the subset of their substrates may vary and their cellular functions may differ significantly. Moreover, the function of PIN

domains may not be limited to exonuclease activity. A sequence similar to PH0500 was found in *P. horikoshii* OT3 itself (PH0708) and two similar ORFs were also found in *P. furiosus* and *P. abyssi* species. The sequence alignment shows that the active-site residues, except Asp98, are not well conserved in the duplicated proteins PH0708 and PF1206 and especially in PAB1741 (Fig. 1). This supports the possibility of their involvement in cellular processes other than those requiring the catalytic activity (*e.g.* protein–protein or protein–nucleic acid interactions). The recent and current structural studies are just the starting points for understanding the potentially diverse role of PIN domains.

This work was supported by a National Project on Protein Structural and Functional Analysis funded by MEXT of Japan (Project PH0500/APCG10117).

References

- Arcus, L. V., Bäckbro, K., Roose, A., Daniel, E. L. & Baker, E. N. (2004). *J. Biol. Chem.* **279**, 16471–16478.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., 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.* **D54**, 905–921.
- Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.
- Clissold, P. M. & Ponting, C. P. (2000). *Curr. Biol.* **10**, 888–890.
- Cohen, G. N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., Poch, O., Prieur, D., Querellou, J., Ripp, R., Thierry, J. C., Van der Oost, J., Weissenbach, J., Zivanovic, Y. & Forterre, P. (2003). *Mol. Microbiol.* **47**, 1495–1512.
- Engl, R. A. & Huber, R. (1991). *Acta Cryst.* **A47**, 392–400.
- Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
- Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
- Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
- Ilyin, A. V., Temple, B., Hu, M., Li, G., Yin, Y., Vachette, P. & Carter, W. C. Jr (2000). *Protein Sci.* **9**, 218–231.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- Janin, J. (1997). *Nature Struct. Biol.* **4**, 973–974.
- Kawarabayasi, Y. *et al.* (1998). *DNA Res.* **5**, 55–76.
- Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
- Laskowski, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Levin, I. *et al.* (2004). *Proteins*, **56**, 404–408.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* **D50**, 869–873.
- Michalopoulos, I., Torrance, G. M., Gilbert, D. R. & Westhead, D. R. (2004). *Nucleic Acids Res.* **32**, D251–D254.
- Navaza, J. (1994). *Acta Cryst.* **A50**, 157–163.
- Nichols, A., Sharp, K. A. & Honig, B. (1991). *Proteins*, **11**, 281–296.
- Noguchi, E., Hayashi, N., Azuma, Y., Seki, T., Nakamura, M., Nakashima, N., Yanagida, M., He, X., Mueller, U., Sazer, S. & Nishimoto, T. (1996). *EMBO J.* **15**, 5595–5605.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Pearl, F. M. G., Lee, D., Bray, J. E., Sillito, I., Todd, A. E., Harrison, A. P., Thornton, J. M. & Orengo, C. A. (2000). *Nucleic Acids Res.* **28**, 277–282.
- Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
- 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.
- Tan, S. & Richmond, T. J. (1998). *Nature (London)*, **391**, 660–666.
- Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* **D55**, 849–861.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
- Waldrop, G. L., Rayment, I. & Holden, H. M. (1994). *Biochemistry*, **33**, 10249–10256.
- Wall, D. & Kaiser, D. (1999). *Mol. Microbiol.* **32**, 1–10.