Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure analysis of PH1161 protein, a transcriptional activator TenA homologue from the hyperthermophilic archaeon *Pyrococcus horikoshii*

Hiroshi Itou,[a]‡ Min Yao,[a,b] Nobuhisa Watanabe[a,b] and Isao Tanaka[a,b]*

[a]Frontier Research Center for Post-genomic Science and Technology, Hokkaido University, Sapporo 001-0021, Japan, and [b]Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

‡ Present address: Biomolecular Structure Laboratory, Structural Biology Center, National Institute of Genetics, Mishima 411-8540, Japan.

Correspondence e-mail: tanaka@castor.sci.hokudai.ac.jp

Received 12 December 2003
Accepted 9 April 2004

PDB Reference: PH1161, 1udd, r1uddsf.

The crystal structure of the *Bacillus subtilis* TenA-homologue protein PH1161 from the hyperthermophilic archaebacterium *Pyrococcus horikoshii* was determined. TenA is known to belong to a new family of activators that stimulate the production of extracellular proteases in *B. subtilis*. A sequence-similarity search revealed that TenA-homologue proteins are widespread in bacteria and archaea, suggesting that this family of proteins plays an essential role in these organisms. In the present study, the first three-dimensional structure of a member of the TenA family of proteins was determined, unexpectedly revealing that the protein has a fold identical to that of haem oxygenase-1. Analysis has also shown that the protein has a unique ligand-binding pocket. Electron density of a bound ligand molecule was observed in this pocket. These results provide a valuable insight into the functional understanding of the TenA family of proteins.

1. Introduction

To adapt to environmental change, organisms possess a wide variety of signal-transduction systems that mediate gene expression. The *Bacillus subtilis deg* system consisting of *degS*, *degU*, *degQ* and *degR* is known to be one such signal-transduction system and regulates the rate of synthesis of degradative enzymes, including an intracellular protease and secreted enzymes (Msadek *et al.*, 1990). The products of the *degS* and *degU* genes, DegS and DegU proteins, have been characterized as being a sensor and a transcriptional regulator, respectively, from their amino-acid sequence similarity (Msadek *et al.*, 1990), and the pair act as a two-component (modulator–effector) pair, similar to the *Escherichia coli* EnvZ–OmpR proteins (Stock *et al.*, 1989).

Another protein factor that affects the expression levels of extracellular protease *via* the *deg* system was identified in *B. subtilis*. This new effector protein, named TenA, is a protein of molecular weight 27.4 kDa that consists of 236 amino acids; it enhances the production of several extracellular proteases by between 11-fold and 55-fold. A functional DegS is required to observe this stimulatory effect, suggesting that TenA acts indirectly to enhance the production of extracellular enzymes (Pang *et al.*, 1991). A BLAST (Altschul *et al.*, 1997) search against all protein-sequence databases revealed that proteins highly homologous to TenA are widely conserved in both the bacteria and archaea, suggesting that these TenA-like proteins play an essential role during gene activation in these organisms. The product of ORF *ph1161* from the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3 (Kawarabayasi *et al.*,

Table 1
Summary of data collection.

Values in parentheses are for the outermost resolution shell.

		SeMAD		
	Native	Peak	Edge	Remote
Wavelength (Å)	1.5418	0.9776	0.9778	0.9815
Resolution (Å)	40–2.15 (2.27–2.15)	40–2.5 (2.59–2.50)		
Space group	$P3_2$	$P222_1$		
No. observed reflections	220886 (31384)	—	—	—
Unique reflections	58735 (8595)	37833 (3496)	37819 (3506)	37695 (3324)
Completeness (%)	99.7 (99.7)	98.6 (93.0)	98.6 (93.4)	98.1 (87.9)
Average redundancy	3.8 (3.7)	6.9 (6.5)	6.9 (6.6)	7.4 (6.7)
Average $I/\sigma(I)$	7.4 (3.0)	17.7 (5.1)	17.1 (4.9)	17.0 (4.8)
$R_{\rm meas}$† (%)	8.2 (28.4)	—	—	—
$R_{\rm merge}$‡ (%)	—	6.8 (24.1)	6.3 (23.9)	6.3 (24.4)
R_λ§ (%)	—	4.9 (10.4)	4.2 (10.3)	—

† $R_{\rm meas} = \sum_h [m/(m-1)]^{1/2} \sum_j |\langle I\rangle_h - I_{h,j}| / \sum_h \sum_j I_{h,j}$, where $\langle I\rangle_h$ is the mean intensity of symmetry-equivalent reflections and m is the redundancy (Diederichs & Karplus, 1997). This factor is also called $R_{\rm r.i.m.}$ (the redundancy-independent merging R factor; Weiss, 2001). ‡ $R_{\rm merge} = \sum_h \sum_j |\langle I\rangle_h - I_{h,j}| / \sum_h \sum_j I_{h,j}$, where $\langle I\rangle_h$ is the mean intensity of symmetry-equivalent reflections. § $R_\lambda = \sum_h \big||F_{\lambda j}| - |F_{\lambda 0}|\big| / \sum_h |F_{\lambda o}|$, where $F_{\lambda j}$ is the structure factor of the data collected at λ_j and $F_{\lambda 0}$ is the structure factor of the data collected at the remote wavelength.

1998) is a protein that has 26% amino-acid sequence identity (45% positives with 5% gaps) to *B. subtilis* TenA. For structural-based functional analysis of TenA, the three-dimensional structure of PH1161 protein, a homologue of TenA, was determined at 2.15 Å resolution. PH1161 is a protein of molecular weight 25.5 kDa that consists of 218 amino acids. The crystal structure shows that the protein has a unique ligand-binding pocket, while exhibiting significant structural similarity to haem oxygenase-1. The structure also showed that an unknown ligand molecule was bound in the pocket. These facts provide valuable information for elucidating the functions of the TenA protein, the detailed biological activities of which are still unknown.

2. Materials and methods

2.1. Sample preparation

The full-length (amino acids 1–218) PH1161 gene was amplified by PCR using the *P. horikoshii* OT3 genome as template DNA and was inserted into a pET-26b vector (Novegen) with *Nde*I and *Xho*I restriction-enzyme sites. The 3′-end primer for this cloning was designed for direct attachment of a 6×His tag to the C-terminal end of the PH1161 gene *via* a two-amino-acid linker Leu and Glu (pET/PH1161-His). The vector was co-transformed into *E. coli* BL21-Star DE3 (Stratagene), the expression host cell with pT-RIL plasmid (Stratagene).

The cells were grown at 310 K. Overexpression of recombinant PH1161-His (termed PH1161 protein) was induced with 1 m*M* IPTG. After IPTG induction, the medium was incubated at 298 K for 14 h with shaking in order to prevent the formation of inclusion bodies of the target protein. The cells were harvested by centrifugation at 3500*g* for 20 min at 277 K and were resuspended in 50 m*M* Tris–HCl pH 7.5 containing 0.15 *M* NaCl, 0.1 m*M* PMSF, 1 m*M* DTT, 5 m*M* $MgSO_4$ and 1 mg DNaseI. The cells were disrupted by a French press (Aminco Inc.) at 83 MPa and the homogenate was clarified by centrifugation at 14 000*g* for 0.5 h at 277 K. The supernatant of the cell extract was heat-treated (333 K) for 0.5 h to remove proteins from the *E. coli* host cell by heat denaturing. After heat treatment, the supernatant was clarified by centrifugation at 14 000*g* for 0.5 h at 293 K. The supernatant was applied onto an Ni-chelating affinity column (Hi-trap Chelating, Amersham Biosciences) equilibrated with binding buffer (20 m*M* Tris–HCl pH 7.5 containing 500 m*M* NaCl and 20 m*M* imidazole). After washing out the unbound proteins with a sufficient amount of binding buffer, PH1161 protein was eluted with elution buffer (20 m*M* Tris–HCl pH 7.5 containing 500 m*M* NaCl and 500 m*M* imidazole). The PH1161-containing fraction was loaded onto a Hi-load 26/60 Superdex200pg size-exclusion chromatography column (Amersham Biosciences). The PH1161 protein was eluted as a single peak. The PH1161 fraction was dialyzed against crystallization buffer (20 m*M* Tris–HCl pH 7.5 containing 200 m*M* NaCl and 10% glycerol) and concentrated by ultrafiltration using Apollo (Orbital Biosciences) to a final concentration of 5 mg ml^{-1}. In this study, the multiple wavelength anomalous diffraction (MAD) method was used for phase calculation. SeMet-substituted PH1161 protein for use in the MAD method was also prepared following the same method as for the native protein with slight modifications.

2.2. Crystallization

Crystallization experiments were performed using the sitting-drop vapour-diffusion method in a 96-well plate (Greiner Bio-one) and the hanging-drop vapour-diffusion method in a 24-well VDX plate (Hampton Research) at 293 K. The initial crystallization condition screening (96-well plate) was carried out using 0.1 ml reservoirs of sparse-matrix crystal screening kits from Hampton Research and deCODE Genetics. Each drop contained 1 µl reservoir solution and an equal volume of protein solution. The best native crystal was obtained using condition No. 31 of Cryo 1 (0.1 *M* citrate pH 5.5, 35% 2-propanol, 5% PEG 1K; deCODE Genetics) and further optimization was carried out using the hanging-drop method. The dimensions of the native crystal were 0.3 × 0.3 × 0.2 mm and the crystal belonged to space group $P3_2$, with unit-cell parameters $a = b = 116.8$, $c = 71.0$ Å, $\gamma = 120^\circ$; the asymmetric unit contained four molecules of PH1161 protein. The crystal volume per protein weight ($V_{\rm M}$; Matthews, 1968) was 2.7 $Å^3$ Da^{-1}, with a solvent content of 55.1%.

The initial SeMet PH1161 crystal was obtained using condition No. 9 of Crystal Screen (0.1 *M* sodium citrate pH 5.6, 0.2 *M* ammonium acetate, 30% PEG 4K; Hampton

Table 2
Summary of phase calculation for SeMet PH1161 protein.

	Remote	Edge	Peak
Resolution (Å)	20–2.5		
Space group	$P222_1$		
MAD phasing			
$R_{\rm Cullis}$†		0.460	0.503
Phasing power_iso‡		3.586	3.105
Phasing power_ano§	0.668	1.162	1.538
FOM¶	0.631		
NCS averaging			
CC††	0.886		

† $R_{\rm Cullis}$ is the mean residual lack-of-closure error divided by the dispersive difference. Values are for centric reflections. ‡ Phasing power_iso is the root-mean-square of F_H/E, where F_H is the dispersive difference of F_H and E is the lack-of-closure error. § Phasing power_ano is as for Phasing power_iso, except that F_H is the anomalous difference of F_H. ¶ FOM is the mean figure of merit. †† CC is the standard linear correlation coefficient between the ρ values of the electron density related by the NCS operation.

Table 3
Final refinement statistics.

Resolution range (Å)	10.0–2.15
No. reflections	58170
Completeness (%)	99.8
Total number of non-H atoms	
Protein	7229
Solvent	155
R factor† (%)	21.8
$R_{\rm free}$ factor‡ (%)	24.6
R.m.s. deviation from standard values	
Bonds (Å)	0.009
Bond angles (°)	1.196
Dihedral angles (°)	18.12
Average *B* factor (Å^2)	34.6
Ramachandran plot§	
Residues in most favoured regions (%)	97.8
Residues in additional allowed regions (%)	1.8
Residues in generously allowed regions(%)	0.4
Residues in disallowed regions (%)	0

† R factor $= \sum |F_{\rm obs} - F_{\rm calc}| / \sum F_{\rm obs}$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure-factor amplitudes. ‡ The $R_{\rm free}$ factor was calculated as for the R factor, using an unrefined subset of reflection data (10%). § The Ramachandran plot was calculated using *PROCHECK* (Laskowski *et al.*, 1993).

Research). Optimization of the crystallization condition was carried out and the best crystal of SeMet PH1161 was obtained using 0.1 *M* sodium citrate pH 5.2, 0.2 *M* ammonium acetate, 30% PEG 4K and 15% glycerol. The dimensions of the SeMet crystal were 0.4 × 0.2 × 0.2 mm and this crystal belonged to space group $P222_1$, with unit-cell parameters $a = 154.1$, $b = 77.6$, $c = 89.6$ Å; the asymmetric unit again contained four molecules of PH1161. The crystal volume per protein weight ($V_{\rm M}$) was 2.6 Å^3 Da^{-1} and the solvent content was 52.7%.

2.3. Data collection

Data-collection statistics are given in Table 1. Native data were collected at a wavelength of 1.5418 Å at 100 K by flash-cooling on an in-house R-AXIS IV^{++} data-collection system equipped with a Micro-Max 7 X-ray generator (Rigaku MSC). The MAD data set was collected at three different wavelengths (0.9778, 0.9776 and 0.9815 Å) from a single SeMet-substituted crystal at 100 K on beamline BL38B1, SPring-8 (Hyogo, Japan). The native data were processed using the programs *MOSFLM* (Leslie, 1993) and *SCALA* (Evans, 1997) and the MAD data were processed using the program *DENZO* (Otwinowski & Minor, 1997).

2.4. Phase calculation

The asymmetric unit contains four molecules of PH1161, each of which contains six SeMet residues. 13 of the 24 Se sites were found using *SOLVE* (Terwilliger & Berendzen, 1999). Initial phases were calculated with the 13 sites and three additional sites were located by difference Fourier methods using *SHARP* (de La Fortelle & Bricogne, 1997). The phases were improved by NCS averaging using *DM* (Cowtan & Main, 1998). The phasing statistics are summarized in Table 2.

2.5. Model building and refinement

After the NCS averaging, a partial model (two-thirds of the molecule) was built using the graphics program *O* (Jones *et al.*, 1991). To improve the electron-density map, multi-crystal NCS averaging using the SeMet crystal and the native crystal was performed using *DMMULTI* (Cowtan & Main, 1998). The native crystal was not isomorphous with the SeMet derivative; the orientation and the position of molecules in the native crystal were therefore determined by a molecular-replacement method using *AMoRe* (Navaza, 1994). From the improved electron-density map after averaging, another part of the partial model was built using *O* at 2.15 Å resolution. The other three molecules in the asymmetric unit were generated by NCS operations from the model. Structure refinement was carried out on the native data using *CNS* (Brünger *et al.*, 1998). Manual model fitting was carried out between refinement rounds using *O*. After some rounds of refinement, the electron density of the unknown ligand molecules that bound to PH1161 were identified from both $2F_{\rm o} - F_{\rm c}$ and $F_{\rm o} - F_{\rm c}$ maps. Water molecules were found using *CNS* and the final model has an *R* factor of 21.8% and a free *R* factor of 24.6% for the data between 10 and 2.15 Å. The final refinement statistics are given in Table 3.

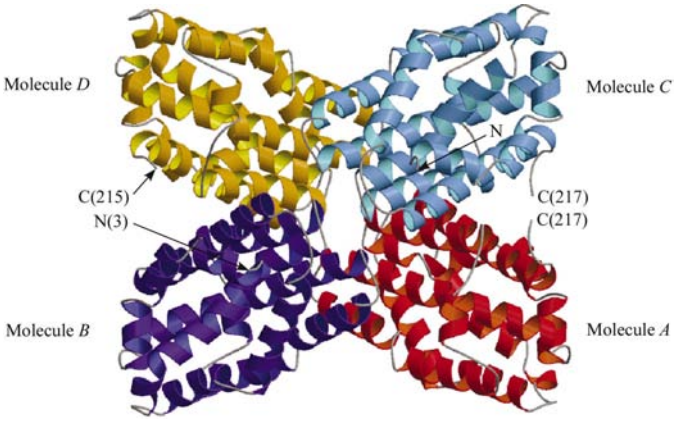


Figure 1
Ribbon representation of the PH1161 protein structure. The protein contains a dimer of dimers in the asymmetric unit, with 222 point-group symmetry. All ribbon representations were generated using *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt & Bacon, 1997).

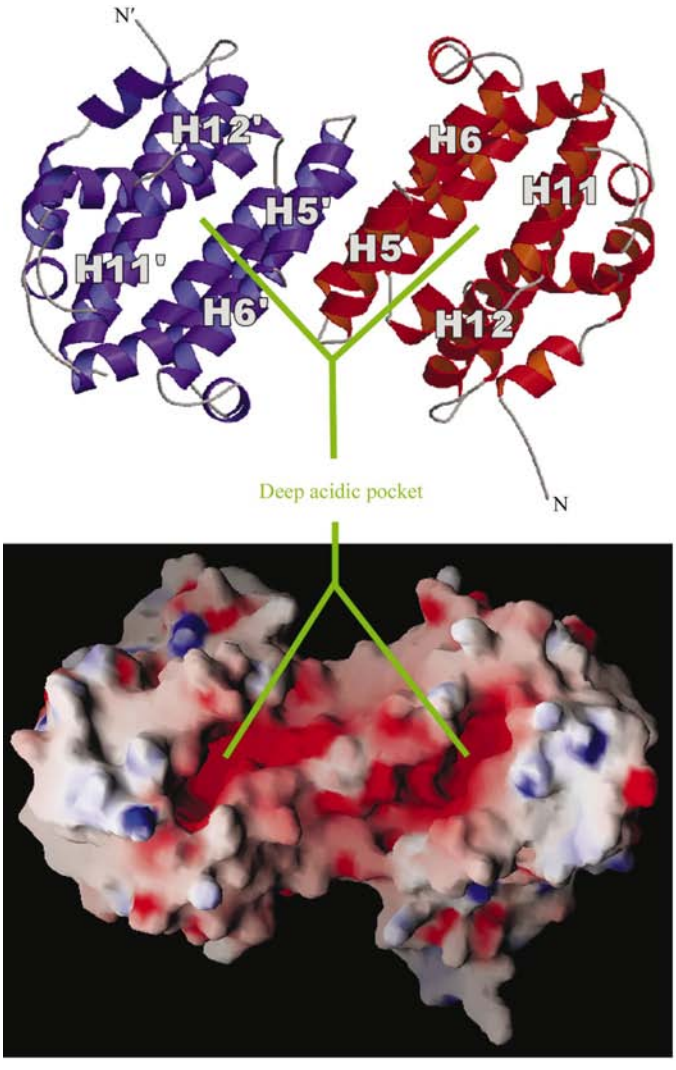
3. Results and discussion

3.1. Structure description: the monomer structure and its oligomer formation

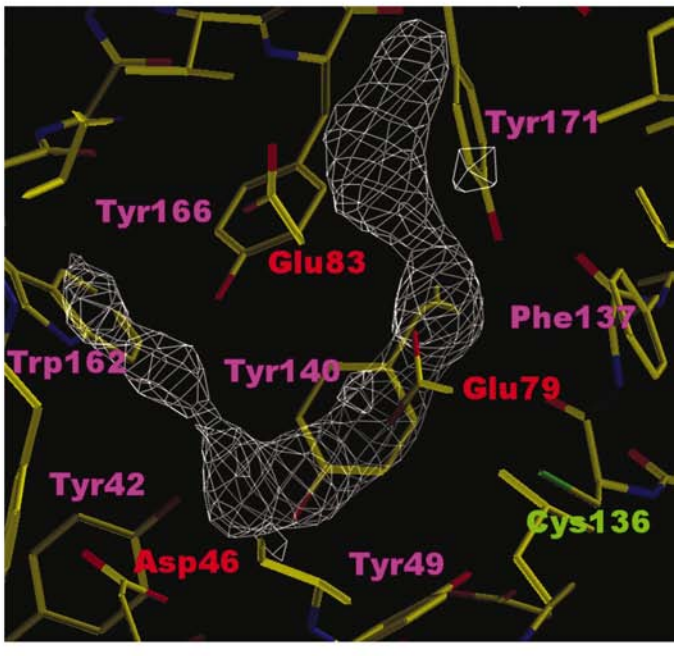
A view of the overall structure of the PH1161 protein is shown in Fig. 1. Atomic models were successfully built from the first residue to Val217 for molecules *A* and *C* and from Val3 to Gly215 for molecules *B* and *D*. The unresolved parts of the models (C-terminal residues, including the attached His tag, and also the N-terminus for molecules *B* and *D*) were a consequence of disorder in the electron-density map of the corresponding regions. PH1161 is an α-helical protein and the molecule consists of 13 α-helices (H1, Ile5–Arg11; H2, Glu14–Phe21; H3, Phe24–Ser31; H4, Leu36–Lys63; H5, Met68–Glu79; H6, Val82–Glu93; H7, Glu98–Lys103; H8, Leu108– Lys123; H9, Ile126–Tyr146; H10, Lys148–Leu151; H11, Lys157–Leu167; H12, Asn169–Asp183; H13, Phe190–Arg213). There is a deep pocket in the area surrounded by helices H5–H6 and H11–H12 (Fig. 2*a*). This pocket is highly charged in the acidic environment and the electron density of an unknown bound ligand molecule was observed in the pocket (Fig. 2*b*).

In Fig. 3, the amino-acid sequences of TenA-homologue proteins are aligned based on the structure of the PH1161 protein. These proteins show 26–77% sequence identity (41–87% positives) with each other and, as shown in this alignment, hydrophobic residues are well conserved among these homologues. The present PH1161 structure shows that most of these conserved hydrophobic residues (coloured yellow) are included in an internal hydrophobic core that maintains the tertiary structure of the monomer molecule. Hydrophobic residues corresponding to Phe21, Leu98, Val109, Leu151 and Ile158 in the PH1161 protein with relatively low levels of conservation are located at the edge of the hydrophobic core formed in the interior of the molecule and amino-acid substitutions in these residues are not likely to cause major structural changes in the protein structure. These findings suggest that the present structure of the PH1161 protein represents the three-dimensional structure of the TenA-homologue proteins.

The native and SeMet-substituted crystals of PH1161 protein used in the present structure analysis contained four molecules in the asymmetric unit. In both crystals, molecules were arranged with 222 point-group symmetry and formed a tetrameric oligomer (Fig. 1). Structural investigation showed that molecules *A* and *C* (and *B* and *D*) interact with each other *via* their H8 and H13 helices. In the interface, hydrophobic



(*a*)



(*b*)

Figure 2
(*a*) Ribbon (top) and surface electric potential (bottom) representations viewed from the direction of the bottom of Fig. 1. In surface electric potential representation, red indicates acidic and blue represents basic charged areas. Each molecule has a deep pocket between helices H5–H6 and H11–H12 and the pockets are highly charged under acidic conditions. All surface electric potential representations were generated using *GRASP* (Nicholls *et al.*, 1991). (*b*) $F_o - F_c$ map for the bound molecule at the acidic pocket of PH1161 protein. The map was contoured at 3σ. Amino acids around the unknown bound ligand molecule are labelled: red, acidic residues; purple, hydrophobic residues; green, conserved Cys residue.

interactions were observed, including two hydrogen-bonding pairs between the residues from each molecule. Each molecule contributes 1359 Å^2 of buried surface area to the formation of this dimer. Dimerization between molecules *A* and *B* (and *C* and *D*) is achieved by an antiparallel interaction between helices H4 and H5 from each molecule. In the interface, 1809 Å^2 of buried surface area from each molecule contributes to the formation of the dimer. The H4 and H5 helices form a relatively flat surface as a dimer interface and extensive hydrophobic interactions and four hydrogen bonds between the residues from these molecules were observed. By this method of dimer formation, the opening of the deep acidic pocket from each molecule is arranged in the same direction and this side of the molecule presents a characteristic functional area of the protein (Fig. 2*a*). These structural features suggest that dimerization of molecules *A* and *B* (and *C* and *D*) is functionally more important. On the other hand, on formation of this tetramer of the PH1161 protein, the side chains of Tyr122 stick out from the molecular surfaces and Val59 residues from each molecule are in close contact, forming a hydrophobic core at the centre of the point-group symmetry. In addition, two hydrogen-bonding pairs between the main-chain carbonyl O atoms of the Tyr122 residues from diagonally opposite molecules (molecules *A* and *D*, and molecules *B* and *C*) *via* a water molecule were observed.

3.2. Functional implications: structural similarities and the bound ligand in the pocket

A DALI structural similarity search (Holm & Sander, 1995) revealed that the PH1161 protein has significant structural similarity to haem oxygenase-1 (HO-1) proteins (human, PDB code 1n45, Schuller *et al.*, 1999; rat, 1dve, Sugishima *et al.*, 2000; bacteria, 1j77; Schuller *et al.*, 2001). In particular, human HO-1 (hHO-1) is most similar to the PH1161 protein (*Z* score = 15.9; r.m.s. difference of 3.5 Å for 214 C$^\alpha$ atoms; Fig. 4).

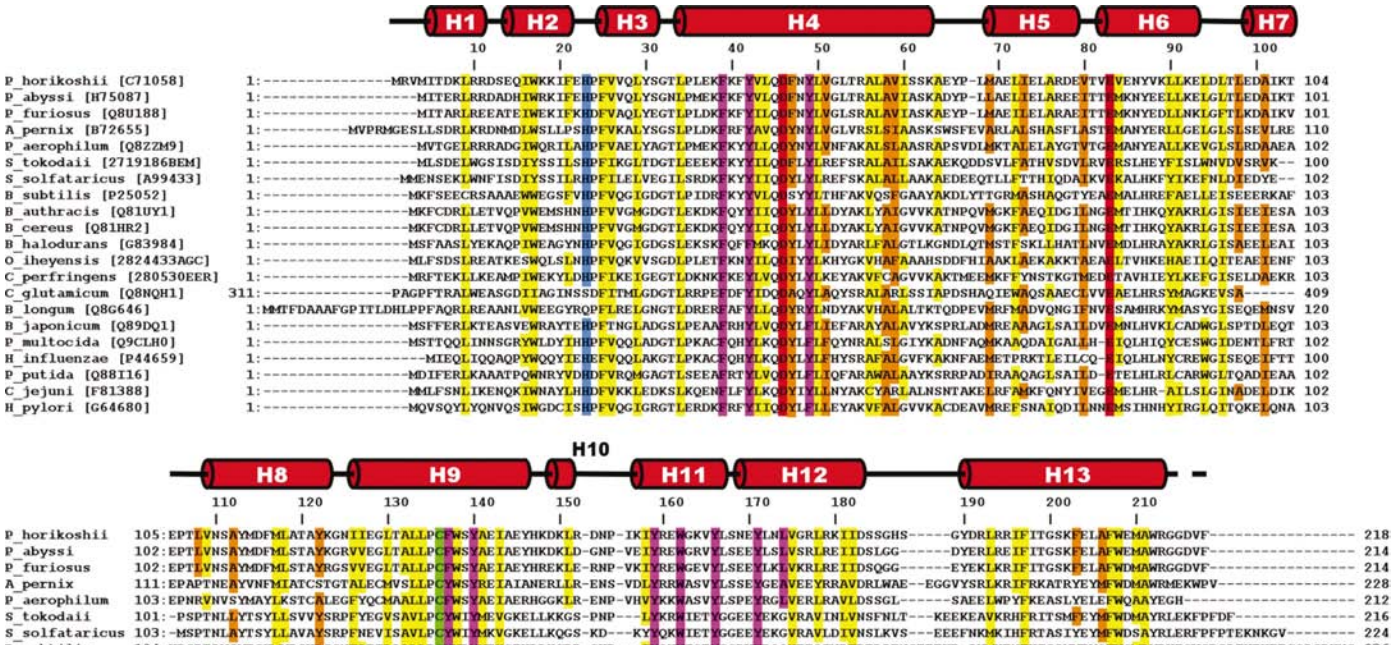


Figure 3
Amino-acid sequence alignment of TenA-homologue proteins. The secondary structure of PH1161 protein is also presented above the alignment. The amino-acid sequence-homology search was performed against all protein-sequence databases *via* the BLAST (Altschul *et al.*, 1997) server with the PH1161 amino-acid sequence and the top 20 retrieved proteins from unique organisms are aligned based on the structure of the PH1161 protein. The accession No. of each protein is also indicated in parentheses after the organism name. *Pyrococcus horikoshii*, *P. abyssi* and *P. furiosus* are euryarchaeota from the archaea. *Aerophyrum pernix*, *Pyrobacterium aerophilum*, *Sulfolobus tokodaii* and *S. solfataricus* are crenarchaeota from the archaea. *Bacillus subtilis*, *B. authracis*, *B. cereus*, *B. halodurans*, *Oceanobacillus iheyensis* and *Clostridium perfringens* are firmicutes from the bacteria. *Corynebacterium glutamicum* and *Bifidobacterium longum* are actinobacteria from the bacteria. *Bradyrhizobium japonicum*, *Pasteurella multocida*, *Haemophillus influenzae*, *Pseudomonas putida*, *Campylobactor jejuni* and *Helicobactor pylori* are proteobacteria from the bacteria. As shown here, TenA-homologue proteins are widely spread in various archaea and bacteria. Amino acids on coloured backgrounds indicate the following: yellow, conserved hydrophobic residues that are involved in the internal hydrophobic core; orange, other conserved hydrophobic residues – most of these residues are involved in intramolecular hydrophobic interactions; blue, conserved His residues; red, acidic residues that construct the ligand-binding pocket; purple, hydrophobic residues that construct the ligand-binding pocket; green, conserved Cys residue in the ligand-binding pocket.

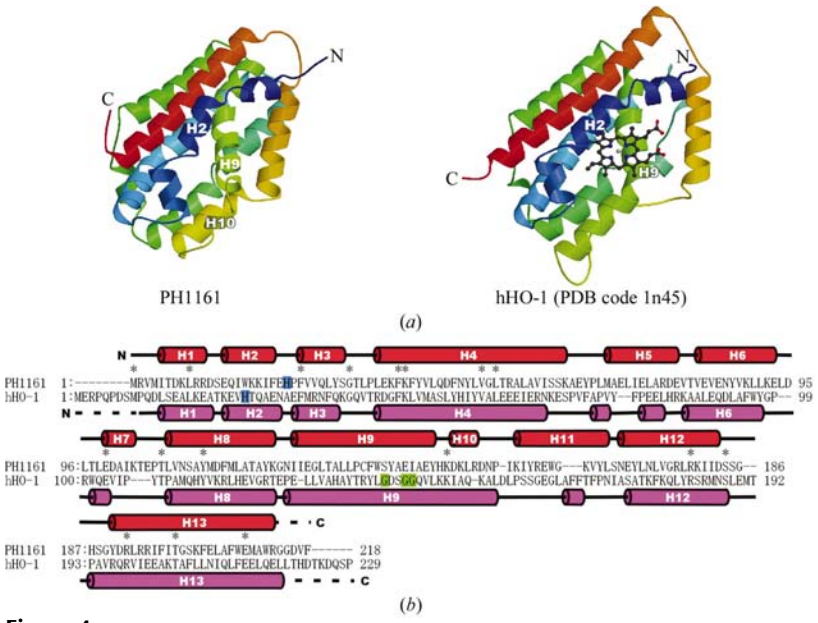


Figure 4
Comparison between PH1161 protein and human HO-1. (*a*) Tertiary structure. The haem molecule bound to hHO-1 is drawn as a ball-and-stick model. (*b*) Primary and secondary structures. The amino-acid sequences of the two proteins are aligned based on their secondary structures. His25 of hHO-1 and His23 of PH1161 are highlighted in blue. Glycine residues in the distal helix of hHO-1 are highlighted in green.

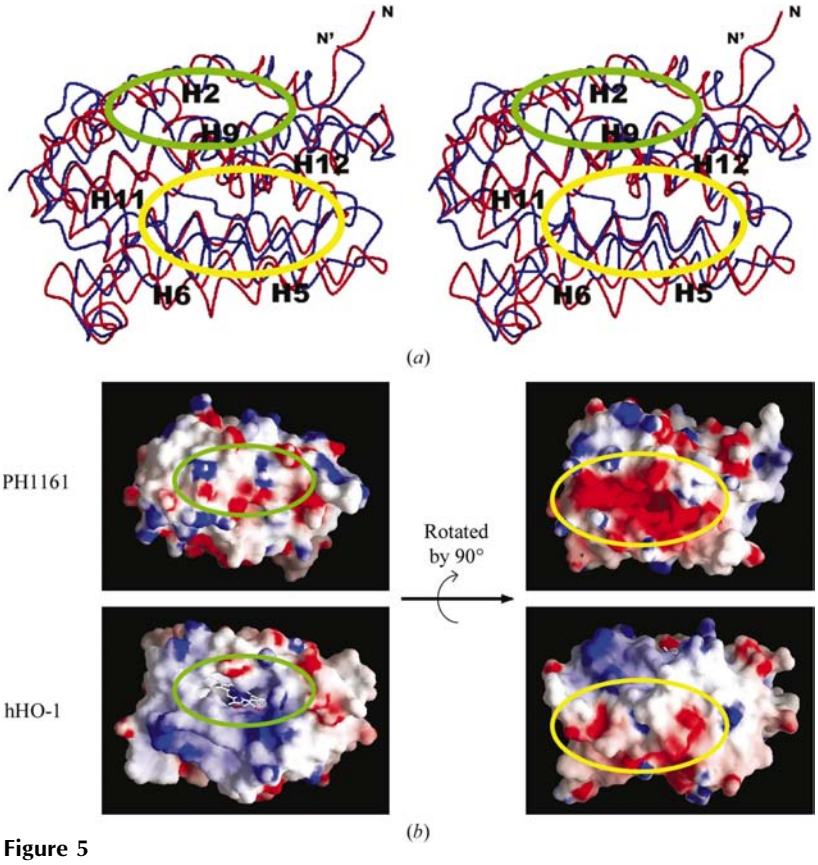


Figure 5
Comparison of the ligand-binding sites of PH1161 protein and hHO-1. The green circle indicates the haem-binding pocket and the yellow circle indicates the acidic ligand-binding pocket. (*a*) Stereoview of the main-chain superposition of PH1161 protein (red) and hHO-1 (blue). Helices constructing their ligand-binding sites are labelled. (*b*) Surface electric charge representation of the PH1161 protein and hHO-1 viewed from same direction. The bound haem molecule is drawn as a ball-and-stick model.

The hHO-1 protein also has a helical structure and its main-chain trace superimposes well onto that of the PH1161 molecule (Fig. 5*a*), although there is only 11% amino-acid sequence identity between them. This extensive structural similarity shows that the PH1161 protein, a TenA-homologue protein, and HO-1 proteins can be classified into the same superfamily. However, although PH1161 and hHO-1 share an identical overall fold, these two proteins exhibit quite different properties. The most notable differences were observed in their ligand-binding sites. HO-1 catalyzes the oxidative cleavage of protohaem to biliverdin (Schuller *et al.*, 1999). There is a haem-binding pocket between its helices H2 (proximal) and H9 (distal) and a ligand-binding His residue is conserved in the proximal helix (Fig. 4). In the middle of the distal helix, Gly residues are also conserved in HO-1 proteins and these glycines play a significant role in the interaction with the haem. The distal helix is kinked at the glycines, providing a structurally favoured orientation for other distal-side residues to interact with the haem. A structural comparison between holo HO-1 and apo HO-1 showed that entry of the haem molecule results in a conformational change in the proximal helix and the latter half of the distal helix (Sugishima *et al.*, 2002). In PH1161, the C^{α}-atom traces of helices H9–H10 differ from those of the HO-1 distal helix H9 (Fig. 5*a*) and the region corresponding to the haem-binding pocket of HO-1 is filled with bulky amino-acid side chains such as Trp, Phe and Arg, with no cleft being observed (Fig. 5*b*). The His residue that is essential for binding to the ferric ion of the haem is also not conserved here. Although the PH1161 protein has a His residue (His23) neighbouring the haem-binding His of HO-1 (Fig. 4), the side chain of His23 is directed toward the opposite side of the helix and the residue no longer has functional importance as a ligand-binding residue, while an amino-acid sequence investigation of the TenA-homologue proteins shows that the histidine is highly conserved among these proteins (Fig. 3). In spite of these differences, the remarkably similar folding of two proteins may suggest these two protein families are derived from a common ancestral protein and that during the evolutionary process these two protein families acquired different functions while retaining similar protein folding and topology.

On the other hand, PH1161 has a distinct deep pocket on the other side of the molecule (Fig. 5*b*). This pocket is formed between helices H5–H6 and H11–H12 (Fig. 2*a*). The HO-1 protein has different main-chain trajectory in the

corresponding region and no such pocket is observed (Fig. 5). As shown in Fig. 2(*b*), this pocket of the PH1161 protein is constructed by the acidic residues Asp and Glu, the hydrophobic residues Phe, Tyr and Trp and a Cys residue. As shown in Figs. 2(*a*) and 5(*b*), this pocket is highly charged in an acidic environment. The acidic character of the pocket arises from the acidic and hydrophobic amino-acid residues clustered here. Sequence alignment shows that the acidic residues Asp46 and Glu83 of the PH1161 protein are highly conserved among TenA-homologue proteins (Fig. 3). The alignment also shows that the hydrophobic residues Tyr42, Tyr49, Phe137, Trp162, Tyr166 and Tyr171 are also highly conserved and that Cys136 is perfectly conserved in TenA-homologue proteins. These findings indicate that this deep acidic pocket of the PH1161 protein is also commonly observed in other TenA-homologue proteins and suggests that TenA proteins may recognize and bind their ligand molecule using this characteristic pocket and exhibit their biological functions here. There is very little literature referring to the TenA protein (Ouzounis & Kyrpides, 1997; Pang *et al.*, 1991) and this limited information is insufficient to reveal how the protein acts. In the present structure analysis, we observed a large electron density from a bound ligand molecule in this pocket in both $F_o - F_c$ and $2F_o - F_c$ maps. The same electron density was found in all molecules in the crystal. As shown in Fig. 2(*b*), the 'sea dragon'-shaped electron density of the bound ligand is sandwiched by the residues that construct the pocket. No compound that fits such an electron-density shape was added during protein purification and crystallization experiments, thus indicating that the PH1161 protein bound this ligand molecule under *in vivo* conditions in the *E. coli* host cell. As described above, amino acids interacting with this ligand molecule are highly conserved in TenA-family proteins, suggesting that these residues are likely to be conserved as functional residues. We could not identify the ligand molecule from the present electron density alone and other experiments using NMR spectroscopy and MS analysis are now in progress in order to characterize this bound ligand.

B. subtilis TenA is a protein that acts indirectly to enhance the production of extracellular proteases (Pang *et al.*, 1991). The tertiary structure of the PH1161 protein revealed that this protein has a unique deep acidic ligand-binding pocket, while the protein shares a common overall structure with another enzyme, haem oxygenase-1, the function of which is quite different. As discussed above, this unique character of the ligand-binding pocket of PH1161 protein suggests that the pocket is the key region for its functional activities. The present results indicate that the ligand-binding pocket of PH1161 protein binds a ligand molecule during its expression in the host cells. Identification and characterization of the unknown ligand molecule bound to the PH1161 protein is likely to accelerate our functional understanding of TenA proteins.

We thank K. Miura of SPring-8 (Hyogo, Japan) for her kind help during data collection, and N. Hirano and K. Hatakeyama of Frontier Research Center for Post-genomic Science and Technology in Hokkaido University for their valuable support during sample preparation and crystallization experiments. We also thank K. Fukuyama of Osaka University for valuable comments on HO-1 protein. This work was supported by a grant-in-aid for the National Project on Protein Structural and Functional Analysis from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3401.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Cowtan, K. & Main, P. (1998). *Acta Cryst.* D**54**, 487–493.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Evans, P. R. (1997). *Proceedings of the CCP4 Study Weekend. Recent Advances In Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 97–102. Warrington: Daresbury Laboratory.
Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kawarabayasi, Y. *et al.* (1998). *DNA Res.* **5**, 55–76.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leslie, A. G. W. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by N. Isaacs, L. Sawyer & S. Bailey, pp. 44–51. Warrington: Daresbury Laboratory.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Msadek, T., Kunst, F., Henner, D., Klier, A., Rapoport, G. & Dedonder, R. (1990). *J. Bacteriol.* **172**, 824–834.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281–296.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Ouzounis, C. A. & Kyrpides, N. C. (1997). *J. Mol. Evol.* **45**, 708–711.
Pang, A. H., Nathoo, S. & Wong, S. (1991). *J. Bacteriol.* **173**, 46–54.
Schuller, D. J., Wilks, A., Ortiz De Montellano, P. R. & Poulous, T. L. (1999). *Nature Struct. Biol.* **6**, 860–867.
Schuller, D. J., Zhu, W., Stojilkovic, I., Wilks, A. & Poulous, T. L. (2001). *Biochemistry*, **40**, 11552–11558.
Stock, J. B., Ninfa, A. J. & Stock, A. M. (1989). *Microbiol. Rev.* **53**, 450–490.
Sugishima, M., Omata, Y., Kakuta, Y., Sakamoto, H., Noguchi, M. & Fukuyama, K. (2000). *FEBS Lett.* **471**, 61–66.
Sugishima, M., Sakamoto, H., Kakuta, Y., Omata, Y., Hayashi, S., Noguchi, M. & Fukuyama, K. (2002). *Biochemistry*, **41**, 7293–7300.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.