Structural analysis of the transcriptional regulator homolog protein from *Pyrococcus horikoshii* OT3

Ui Okada, Naoki Sakai, Min Yao, Nobuhisa Watanabe, and Isao Tanaka
Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo, Japan

Introduction

PH1061 is a hypothetical protein of 100 residues (11.4 kDa) from the hyperthermophilic archaebacterium, *Pyrococcus horikoshii* OT3[1], and is conserved among many archaeal and bacterial organisms. Although the functions in these organisms are still unknown, a Pfam[2] database search revealed that PH1061 has a helix-turn-helix (HTH) motif ("HTH_5 motif" in the Pfam database), and suggested that it is a DNA-binding protein, like members of the ArsR family[3]. ArsR is a transcriptional repressor of the arsenic resistance operon in *Escherichia coli* and other members of this family involve zinc-sensing transcriptional repressors, such as CzrA[4] from *Staphylococcus aureus* and SmtB[5] from *Synechococcus* pcc7942. The primary sequence similarities to PH1061 are 20%, 21%, and 6.6% with ArsR, CzrA, and SmtB, respectively. There are 10 proteins with the HTH_5 motif in *P. horikoshii*, but none of their functions are known. For structure-based functional analysis, the three-dimensional structure of PH1061 was determined at a resolution of 2.05 Å using the multi-wavelength anomalous diffraction (MAD) method.

Materials and Methods

The PH1061 gene was amplified by polymerase chain reaction (PCR) from *P. horikoshii* OT3 genomic DNA. The PCR product was cloned into the pET-22b (+) vector (Novagen). PH1061 protein was expressed in *E. coli* strain B834 (DE3) as a C-terminal histidine-tagged fusion protein. Cells were grown in Luria-Bertani (LB) medium at 37°C and induced with 0.5 m*M* isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were harvested by centrifugation, washed with buffer A (50 m*M* sodium phosphate pH 6.0, 300 m*M* NaCl), and disrupted using a French press. Cell debris was removed by centrifugation, and the supernatant was incubated at 70°C for 30 min, then denatured proteins from host *E. coli* cells were removed by centrifugation. The extract was mixed to with Ni-nitrilo triacetic acid (Ni-NTA) resin (QIAGEN), washed with buffer A containing 20 m*M* imidazole, and the bound proteins were eluted by adding 500 m*M* imidazole. The eluate was dialyzed against buffer A, and loaded onto a HiLoad 26/60 Superdex 75pg column (Amersham Bioscience) equilibrated with buffer A for size exclusion chromatography. Fractions containing PH1061 were pooled and applied

to a POROS QE/M column (Applied Biosystems) equilibrated with buffer A. PH1061 passed through the column and contaminants were adsorbed. The flow-through fractions were pooled, buffer A was exchanged for buffer B (20 m*M* 2-(N-morpholino) ethanesulfonic acid (MES) pH 6.0, 200 m*M* NaCl), and concentrated to 25 mg ml^{-1}. Selenomethionine (SeMet)-labeled PH1061 was expressed in *E.coli* strain B834 (DE3) grown in minimal medium containing Se-Met, and purified as described for the native protein.

Native and Se-Met PH1061 crystals were grown at 20°C by the vapor diffusion method in hanging drops. The crystallization conditions for PH1061 consisted of 100 m*M* Tris pH 8.7, and 1.5 *M* diammonium hydrogen phosphate. X-ray diffraction data of Se-Met PH1061 were collected to 2.05 Å resolution at 100 K at the BL41XU in SPring-8 (Hyogo, Japan). The diffraction images were processed using the HKL 2000 package[6]. The Se-Met PH1061 crystal belongs to the space group $P4_32_12$ with unit cell dimensions of $a = b = 45.2$ Å, $c = 96.2$ Å. Statistics for this data are shown in Table I. The structure of PH1061 was determined by the MAD method. Determination of Se sites and phase calculations were carried out with the program SOLVE[7]. The initial phases were improved further with RESOLVE[8]. The model of PH1061 was built automatically using ARP/wARP[9], and optimized manually using the program O[10]. The molecular coordinates were refined, and water molecules were located using the program CNS[11]. The atomic coordinates of PH1061 have been deposited in the PDB[12] with the accession code 1UB9.

Results and Discussion

The final model of PH1061 contains all residues 1–100, 57 molecules of water, and has *R* and *R*-free factors of 21.6 and 24.2%, respectively. PH1061 is an α+β structure consisting of 6 α-helices and 3 β-strands (Fig. 1(a)). Two of the helices (α4 (33-40), α5 (43–57)) and two β-strands (β2 (60–65), β3 (71–76)) form a winged-HTH motif seen in many bacterial transcription regulatory proteins, and helix α5 corresponds to the recognition helix. The results of structural similarity analysis using the DALI[13] server indicated that the structure of PH1061 resembles those of some repressor proteins, such as MarR[14], which is a regulatory protein of multidrug resistance (MDR)[15,16] in *E. coli*, with 2.5 Å root mean square distance (r.m.s.d.) similarity per 94% of Cα, and zinc-sensing repressor SmtB with 2.2 Å r.m.s.d. similarity per 81% of Cα, although these molecules have low levels of amino acid sequence homology of 16% and 6.6%, respectively. These observations suggest that PH1061 is a DNA-binding protein.

PH1061 forms a tightly associated dimer related by a crystallographic twofold rotation (Fig. 1(b)). Hydrophobic residues from α1 (1–9), α3 (15–29), and α6 (77–99)

are buried at the dimerization interface. The dimerization reduces the percentage of the hydrophobic surface area from 28.8% in the monomer to 18.6% in the dimer. The results of dynamic light scattering and size exclusion chromatography also indicated that PH1061 forms a dimer in solution. These observations suggest that the functional state of PH1061 is a dimer.

In the dimer of PH1061, there is a positively charged groove between α5 helices of each monomer (Fig. 1(c)). The surface of this groove is formed with positively charged side-chain atoms of residues His11, Arg18, His50, and Arg52 in each monomer, and these residues are conserved among the proteins homologous to PH1061 (Fig. 1(d)). The surface of the wings consisting of β2 and β3 are also positively charged with conserved residues Lys61, Lys64, and Arg71. This positively charged surface is adequate for interaction with DNA.

Generally, HTH motif proteins interact with the target DNA as a homodimer by the penetration of recognition helices from each monomer into the major grooves. In this case, the theoretical distance between recognition helices is 34 Å, which corresponds with the pitch of linear B-DNA. However, the distance between the two α5 recognition helices of PH1061 dimer is approximately 18 Å, and it is too short to make the conventional interaction with DNA. The systematic evolution of ligands by exponential enrichment (SELEX)[17,18] analysis using PH1061 dimers revealed no specific binding site in the genome of *P. horikoshii*. It is possible that the dimer conformation observed in this crystal structure is a closed conformation that cannot bind DNA, and some effector(s) may induce a conformational change in the dimer widening the distance between recognition helices to allow interaction with DNA.

Acknowledgments

We would like to thank M. Kawamoto, and H. Sakai of the Japan Synchrotron Radiation Research Institute (JASRI) for their kind assistance with the X-ray diffraction data collection on the beamline BL41XU at SPring-8. This work was supported by a research grant from the National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

Fig. 1 The structure of PH1061. (**a**) Ribbon representation of PH1061 monomer. The color-coding is from blue (N-terminus) to red (C-terminus). (**b**) Dimer structure of PH1061. **i**), **iii**) Ribbon representation with the monomers colored blue and red. **ii**), **iv**) Electrostatic surface representation colored according to surface potential (blue=positive; red=negative). The same viewing angle as in i) and ii) or iii) and iv). (**c**) Close-up view of the groove between α5 recognition helices of PH1061 dimer. Positively charged side chains of conserved residues are shown as stick models colored as follows: carbon=green; nitrogen=blue. Figures (a–c) were generated using PyMOL (Delano Scientific LLC). (**d**) Multiple sequence alignment of PH1061 with homologous proteins from *Pyrococcus furiosus*, *Sulfolobus tokodaii*, *Halovacterium sp.*, *Methanococcus jannaschii*, *Caulobacter crescentus*, *Mesorhizobium loti*, and *Bradyrhizobium japonicum*. Sequences were aligned with CLUSTALW. The secondary structure elements of PH1061 are illustrated as tubes for α-helices and arrows for β-strands. Numbering is given according to the primary sequence of PH1061. Completely conserved residues are highlighted in yellow, strongly similar residues are highlighted in light green, and weakly similar residues are highlighted in light blue.

Table 1. Summary of data collection and refinement statistics

	Peak	Edge	Remote
Wavelength (Å)	0.9794	0.9796	0.9000
Unit cell (Å)		$a = b = 45.2$, $c = 96.2$	
Space group		$P4_32_12$	
Resolution (Å)	50-2.05	50-2.05	50-2.05
Completeness (%)	99.8	99.9	99.5
Unique reflections	6,749	6,752	6,696
Averaged redundancy	13.5	13.4	13.7
[a]R_{sym} (%)	8.9	6.7	6.0
No. of nonhydrogen atoms			799
No. of water molecules			57
R / *R*-free factor (%)			21.6 / 24.2
RMS deviation in bond length (Å)			0.007
RMS deviation in bond angle (degree)			1.22

[a]$R_{sym} = \Sigma_h \Sigma_j | I_{h,j} - \langle I \rangle_h | / \Sigma_h \Sigma_j I_{h,j}$, where $I_{h,j}$ is observed intensity and $\langle I \rangle_h$ is average intensity obtained from multiple observations of symmetry-related refractions.

1. Kawarabayasi Y, Sawada M, Horikawa H, Haikawa Y, Hino Y, Yamamoto S, Sekine M, Baba S, Kosugi H, Hosoyama A, Nagai Y, Sakai M, Ogura K, Otsuka R, Nakazawa H, Takamiya M, Ohfuku Y, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Kikuchi H. Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, Pyrococcus horikoshii OT3. DNA Res 1998;5(2):55-76.
2. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. Nucleic Acids Res 2004;32(Database issue):D138-141.
3. Busenlehner LS, Pennella MA, Giedroc DP. The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. FEMS Microbiol Rev 2003;27(2-3):131-143.
4. Eicken C, Pennella MA, Chen X, Koshlap KM, VanZile ML, Sacchettini JC, Giedroc DP. A metal-ligand-mediated intersubunit allosteric switch in related SmtB/ArsR zinc sensor proteins. J Mol Biol 2003;333(4):683-695.
5. Cook WJ, Kar SR, Taylor KB, Hall LM. Crystal structure of the cyanobacterial metallothionein repressor SmtB: a model for metalloregulatory proteins. J Mol Biol 1998;275(2):337-346.
6. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology 1997;276:307-326.
7. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr 1999;55 (Pt 4):849-861.
8. Terwilliger TC. Reciprocal-space solvent flattening. Acta Crystallogr D Biol Crystallogr 1999;55(11):1863-1871.
9. Lamzin VS, Wilson KS. Automated refinement of protein models. Acta Crystallogr D Biol Crystallogr 1993;49(Pt 1):129-147.
10. Jones TA, Zou JY, Cowan SW, Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 1991;47 (Pt 2):110-119.
11. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54 (Pt 5):905-921.
12. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28(1):235-242.
13. Holm L, Sander C. Touring protein fold space with Dali/FSSP. Nucleic Acids Res 1998;26(1):316-319.
14. Alekshun MN, Levy SB, Mealy TR, Seaton BA, Head JF. The crystal structure of MarR, a

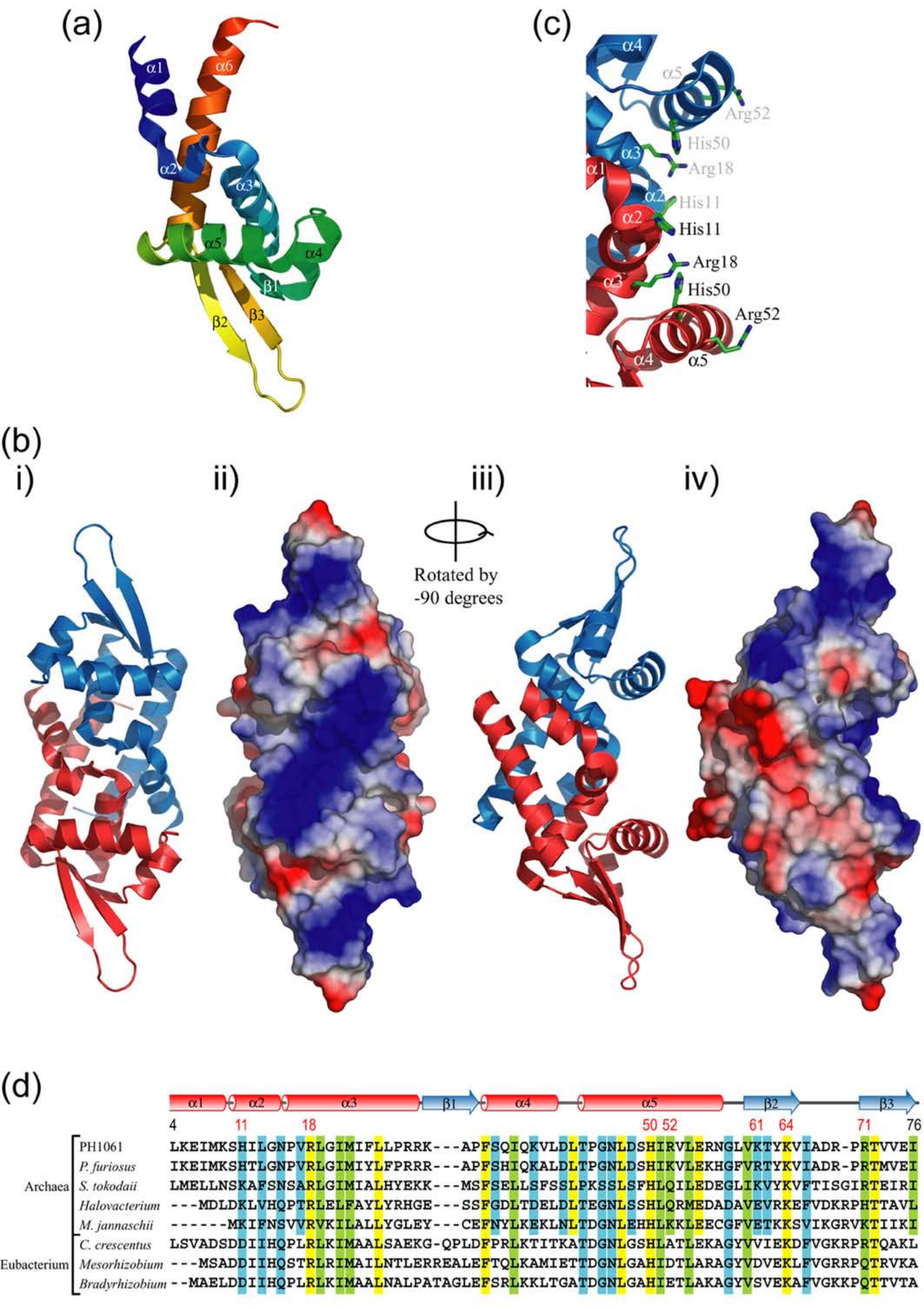
regulator of multiple antibiotic resistance, at 2.3 A resolution. Nat Struct Biol 2001;8(8):710-714.

15. Grkovic S, Brown MH, Skurray RA. Regulation of bacterial drug export systems. Microbiol Mol Biol Rev 2002;66(4):671-701, table of contents.

16. Schumacher MA, Brennan RG. Structural mechanisms of multidrug recognition and regulation by bacterial multidrug transcription factors. Mol Microbiol 2002;45(4):885-893.

17. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 1990;249(4968):505-510.

18. Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. Nature 1990;346(6287):818-822.


(a)
α1
α6
α2
α3
α5
α4
β1
β2
β3
(c)
α4
α5
Arg52
His50
α3
Arg18
α1
α2
His11
α2
His11
Arg18
α3
His50
Arg52
α4
α5
(b)
i)
ii)
iii)
iv)
Rotated by
-90 degrees
(d)
α1
α2
α3
β1
α4
α5
β2
β3
4
11
18
50 52
61
64
71
76
PH1061 LKEIMKSHILGNPVRLGIMIFLLPRRK---APFSQIQKVLDLTPGNLDSHIRVLERNGLVKTYKVIADR-PRTVVEI
P. furiosus IKEIMKSHTLGNPVRLGIMIYLFPRRR---APFSHIQKALDLTPGNLDSHIKVLEKHGFVRTYKVIADR-PRTMVEI
Archaea S. tokodaii LMELLNSKAFSNSARLGIMIALHYEKK---MSFSELLSFSSLPKSSLSFHLQILEDEGLIKVYKVFTISGIRTEIRI
Halovacterium ---MDLDKLVHQPTRLELFAYLYRHGE---SSFGDLTDELDLTEGNLSSHLQRMEDADAVEVRKEFVDKRPHTTAVL
M. jannaschii ------MKIFNSVVRVKILALLYGLEY---CEFNYLKEKLNLTDGNLEHHLKKLEECGFVETKKSVIKGRVKTIIKI
C. crescentus LSVADSDDIIHQPLRLKIMAALSAEKG-QPLDFPRLKTITKATDGNLGSHLATLEKAGYVVIEKDFVGKRPRTQAKL
Eubacterium Mesorhizobium ---MSADDIIHQSTRLRIMAILNTLERREALEFTQLKAMIETTDGNLGAHIDTLARAGYVDVEKLFVGRRPQTRVKA
Bradyrhizobium --MAELDDIIHQPLRLKIMAALNALPATAGLEFSRLKKLTGATDGNLGAHIETLAKAGYVSVEKAFVGKKPQTTVTA