Crystal structure of the PH1932 protein, a unique archaeal ArsR type winged-HTH transcription factor from Pyrococcus horikoshii OT3

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · MARCH 2007

Impact Factor: 2.63 · DOI: 10.1002/prot.21851 · Source: PubMed

4 AUTHORS, INCLUDING:

Hiroshi Itou
Chiome Bioscience Inc.
21 PUBLICATIONS **319** CITATIONS
SEE PROFILE

Min Yao
Hokkaido University
188 PUBLICATIONS **3,510** CITATIONS
SEE PROFILE

Nobuhisa Watanabe
Nagoya University
116 PUBLICATIONS **1,130** CITATIONS
SEE PROFILE

WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

Crystal structure of the PH1932 protein, a unique archaeal ArsR type winged-HTH transcription factor from *Pyrococcus horikoshii* OT3

Hiroshi Itou,[1] Min Yao,[1,2] Nobuhisa Watanabe,[1,2] and Isao Tanaka[1,2]*

[1] Frontier Research Center for Post-genomic Science and Technology, Hokkaido University, Sapporo 001-0021, Japan

[2] Faculty of Advanced Life Sciences, Hokkaido University, Sapporo 060-0810, Japan

***Key words:** ArsR homolog protein; crystal structure analysis; Pyrococcus horikoshii PH1932 protein; structural genomics; winged helix-turn-helix motif.*

INTRODUCTION

PH1932 is a protein of 192 amino acids, which is the product of the ORF *ph1932* from the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3.[1] Amino acid sequence analysis suggested that this protein is composed of two domains. The N-terminal domain is expected to contain a winged helix-turn-helix (HTH) motif similar to that from a bacterial metal-sensing transcription factor, ArsR protein, while no structural or functional information is available for the C-terminal domain. ArsR homologs are widespread in bacteria, and form a class of prokaryotic metal-sensor proteins named the SmtB/ArsR family.[2] These metalloregulatory transcription factors repress expression of operons linked to detoxification of several toxic metal ions. Therefore, their derepression results in enhancement of bacterial tolerance for survival under harsher environments. Although proteins homologous to the PH1932 protein are found in various archaeal species, their biological functions have yet to be characterized. We report here the crystal structure of the PH1932 protein determined by X-ray crystallography at 2.5-Å resolution.

MATERIALS AND METHODS

Sample preparation and crystallization experiments of the full-length PH1932 protein were performed essentially as described previously.[3] In this study, the best protein crystals were obtained using 0.1*M* Na-HEPES (pH 6.8), 0.1*M* $MgCl_2$, 10% PEG400, and 10% glycerol. The dimensions of the crystal were 0.2 × 0.2 × 0.3 mm^3, and the crystal belonged to the space group *R*32 with unit cell parameters $a = b = 90.6$ Å, $c = 132.3$ Å, and $\gamma = 120^\circ$. The MAD data were collected at three different wavelengths from a single crystal of the PH1932 protein at 100 K on the beamline BL41XU, SPring-8 (Hyogo, Japan), and processed using the program HKL2000.[4] The data collection statistics are given in Table I. Selenium sites were found using the program SOLVE,[5] and phasing and subsequent phase improvement by solvent flipping were performed using SHARP-SOLOMON.[6] Model building was performed automatically using the program RESOLVE,[7,8] and this trial succeeded in building up to 80% of the total protein subunit including side chains. The remaining parts of the model were built manually using the graphic program O.[9] Structure refinement was carried out semiautomatically using the program LAFIRE[10,11] with CNS.[12] The final model had an *R*- factor of 20.4% and a free *R*-factor of 26.8% for the

Hiroshi Itou's current address is Biomolecular Structure Laboratory, Structural Biology Center, National Institute of Genetics, Mishima 411-8540, Japan.
Grant sponsor: Ministry of Education, Culture, Sports, Science, and Technology of Japan (National Project on Protein Structural and Functional Analysis).
*Correspondence to: Isao Tanaka, Faculty of Advanced Life Sciences, Hokkaido University, Sapporo 060-0810, Japan. E-mail: tanaka@castor.sci.hokudai.ac.jp
Received 19 June 2007; Revised 26 September 2007; Accepted 1 October 2007
Published online 12 December 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21851

Table I
The Summary of Data Collection, Phasing, and Refinement Statistics

	Se-MAD		
	Peak	Edge	Remote
Data collection			
Wavelength (Å)	0.9792	0.9795	0.9685
Space group		*R*32	
Resolution (Å)		40–2.6 (2.69–2.60)	40–2.5 (2.59–2.50)
Unique reflections	6641 (645)	6646 (647)	7284 (699)
Completeness (%)	99.9 (99.4)	100 (100)	99.4 (99.4)
Averaged redundancy	10.7 (8.3)	10.8 (9.0)	10.5 (5.8)
Averaged $I/\sigma(I)$	6.5	6.9	8.0
R_{merge}[a] (%)	11.4 (36.5)	8.8 (33.2)	8.3 (29.5)
R_{λ}[b] (%)	5.0 (12.0)	4.9 (10.2)	—
MAD phasing			
Resolution range (Å)		20–2.6	
R_{Cullis}[c]	0.531	0.442	
Phasing power_iso[d]	3.002	3.436	
Phasing power_ano[e]	2.916	1.387	2.565
FOM[f]			0.633
Solvent flipping (SOLOMON)			
FOM			0.9554
CC[g]			0.9952
Refinement			
No. of nonhydrogen atoms			1582
Water			25
R^h/*R*-free[i] (%)			20.4/26.8
RMSD bond-lengths (Å)			0.004
RMSD bond-angles (deg.)			0.937
Average *B*-factor (Å^2)			28.12
Ramachandran plot[j]			
In most favored regions (%)			96.1
In additional allowed regions (%)			3.9

Values in parentheses are for the outermost resolution shell.
[a] $R_{merge} = \Sigma_h\Sigma_j |\langle I\rangle_h - I_{h,j}|/\Sigma_h\Sigma_j I_{h,j}$, where $\langle I\rangle_h$ is the mean intensity of symmetry-equivalent reflections.
[b] $R_{\lambda} = \Sigma_h||F_{\lambda\ j}| - |F_{\lambda,o}||/\Sigma_h |F_{\lambda,o}|$, where $F_{\lambda\ j}$ is the structure factor of the data collected at λj and $F_{\lambda,o}$ is the structure factor of the data collected at the remote wavelength.
[c] R_{Cullis} is the mean residual lack-of-closure error divided by the dispersive difference. Values are for centric reflections.
[d] 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.
[e] Phasing power_ano is as for phasing power_iso, except that F_H is the anomalous difference of F_H.
[f] FOM is the mean figure of merit.
[g] CC is the standard linear correlation coefficient between observed and calculated structure factor amplitudes.
[h] $R = \Sigma\ |F_{obs} - F_{cal}|/\Sigma F_{obs}$, where F_{obs} and F_{cal} are observed and calculated structure factor amplitudes.
[i] *R*-free value was calculated for the *R*-factor, using an unrefined subset of reflections data (10%).
[j] The Ramachandran plot was calculated by PROCHECK.[13]

data between 10-Å and 2.5-Å resolution. The phasing and final refinement statistics are also given in Table I.

RESULTS AND DISCUSSION

The final model of the PH1932 protein includes Ala2 to Lys191, and 25 water molecules (PDB code: 1uly). The crystal contains one PH1932 protein subunit per asymmetric unit, and this subunit forms a homodimer with another subunit related by the crystallographic symmetry. The PH1932 protein subunit consists of two domains [Fig. 1(a)]. The N-terminal domain has an α/β architecture containing a four-helix bundle (α1–α4) flanked by two pairs of antiparallel β-sheets (β1–β5, and β2, β3–β4). The C-terminal domain consists of four α-helices (α5–α8) and one 3_{10} helix, and two subunits are dimerized via this C-terminal domain. The N-terminal domain has a winged- HTH motif (right-handed three-helix bundle from α2 to α4 and twisted long β-sheets β2, β3–β4). This motif is one of the most common prokaryotic DNA-binding motifs and is also found in some eukaryotic DNA-binding proteins such as histone-H5[20] or HNF-3γ.[21] DALI structure similarity search[16] showed that the PH1932 protein had a number of structural neighbors with its N-terminal domain in various nucleotide-binding proteins, including those of eukaryotes, such as human ADAR1[22] or yeast ESCRT-II.[23] The closest neighbors were found in bacteria [Fig. 1(b)]: *Synechococcus* sp. SmtB[17] (PDB 1smt, *z*-score = 10.9, root mean

DOI 10.1002/prot

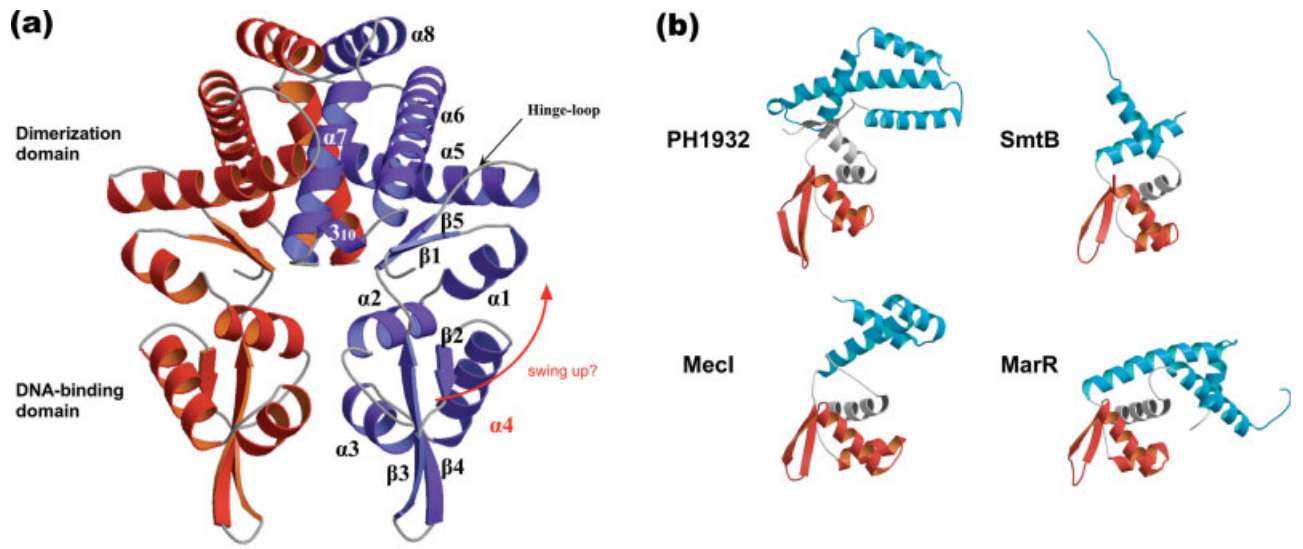

Figure 1

*(**a**) Ribbon representation of the PH1932 protein dimer. The subunits are colored blue and red, respectively. The biological unit of this protein was also supposed to be a dimer based on the estimated molecular weight using size exclusion chromatography (SEC). The second helix in the winged-HTH motif, a putative DNA-recognition helix, is labeled red. Figure 1 was generated using MOLSCRIPT[14] and RASTER3D.[15] (**b**) Structural neighbors of the PH1932 protein found by DALI search.[16] The winged-HTH motif in each subunit is colored red. The structures of their dimerization domain are dissimilar reflecting their divergent ligands (colored sky blue). SmtB[17] and MecI[18] are transcription factors responsive to Zn^{2+} ions and the antibiotic methicillin, respectively. MarR[19] is a transcription factor responsive to a wide range of compounds, including antibiotics and organic solvents.*

square deviation (r.m.s.d.) = 3.1 Å, on 79 Cα atoms of 122 amino acids), *Staphylococcus aureus* MecI[18] (PDB 1okr, z = 9.4, r.m.s.d. = 4.9 Å on 91 Cα of 120 a.a.), BlaI[24] (PDB 1xsd, z = 8.4, r.m.s.d. = 4.8 Å on 95 Cα of 125 a.a.), *Thermotoga maritima* IclR[25] (PDB 1mkm, z = 8.1, r.m.s.d. = 4.9 Å on 83 Cα of 246 a.a.), and *Escherichia coli* MarR[19] (PDB 1igs, z = 5.8, r.m.s.d. = 5.8 Å on 88 Cα of 138 a.a.). All of these proteins are transcription factors involved in environmental adaptation systems of bacteria, suggesting that PH1932 protein maybe a transcriptional regulator of similar function. Although it is difficult to identify the ligand that controls the function of the PH1932 protein as a transcription factor based only on its structural information, its three-dimensional structure provides some clues to the functions of this protein: that is, ligand-binding and DNA-binding. PH1932 protein is homodimeric with its C-terminal domain, and by this dimerization, a deep cavity is formed in the interior side of the domains. The cavity is highly acidic, and traps two water molecules (Fig. 2). This suggests a possible role of the cavity as a ligand-binding site. In this putative ligand-binding site, the OE atom of Glu135 makes hydrogen bond with the water molecule. The residues Glu166′, Thr170′, and Tyr179 also exist in position that would allow interactions with a ligand or an ion instead of the water molecule in a cooperative manner with residues Glu135. Another water molecule is trapped by the NE atom of Trp167′, and other residues Arg104, Lys108, and His163′ that are likely to take part in ligand binding also exist around this second water-binding site. These residues are candidates for the effec-

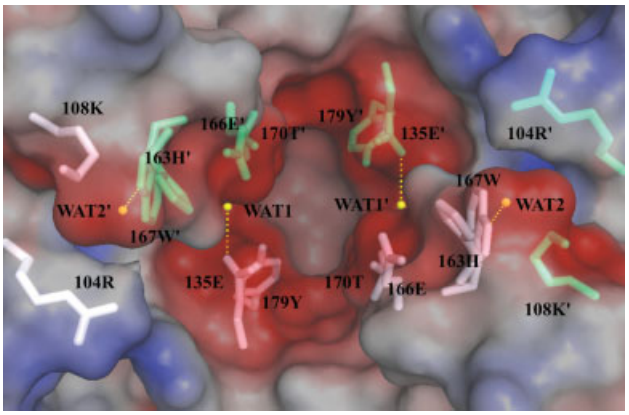

Figure 2

Close-up molecular surface representation around the putative ligand-binding pocket observed on the inside of the dimerization domains. Red and blue on the molecular surface represent negatively or positively charged potentials, respectively. As shown here, the pocket has a significant negative charge. Bound water molecules are shown as yellow balls, and candidate side chains for ligand binding are represented as white stick models. Symmetry-related mates are also indicated as lime-green sticks with dashed labels. Each yellow dotted line shows hydrogen bond between the water and the ligated side chain. This figure was generated using PyMOL (DeLano Scientific LLC).

tor binding residues involved in the ligand–protein interactions. On the other hand, a pair of putative N-terminal DNA-binding domains face each other in the present structure. The putative DNA-recognition helices are located side-by-side on both sides of the dimer axis. This orientation is not suited for binding to straight DNA. The possibility of the protein binding to bent DNA, similar to the integration host factor protein[26] cannot be excluded, but it is more likely that the present structure represents a closed form of the protein dissociated from DNA. Both domains in the subunit are connected with the linker region between β4 and α5, and this will allow reorientation of each DNA-binding domain to bind DNA; each DNA-binding domain in a dimer must be swung up to locate its DNA-recognition helix in a suitable position to interact with target DNA [Fig. 1(a)].

The tertiary structure of the PH1932 protein indicated that this protein subunit has a winged-HTH motif structurally similar to those of bacterial transcription factors known to act as transcriptional repressors in regulating resistance systems against cytotoxic compounds, such as antibiotics or heavy metal ions. The C-terminal domain of the PH1932 protein is responsible for dimerization and forms a unique hat-shaped helix-bundle. The inside of the hat is a possible effector-binding site of this protein. These results suggest that the PH1932 protein may be an archaeal transcription factor involved in environmental adaptation.

ACKNOWLEDGMENTS

We thank SPring-8 BL41XU beamline staff and the technical assistants of the Frontier Research Center for Post-genomic Science and Technology, Hokkaido University, for their valuable support.

REFERENCES

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, Takimiya 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 hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. DNA Res 1998;5:55–76.
2. 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: 131–143.
3. Itou H, Yao M, Watanabe N, Tanaka I. Structure analysis of PH1161 protein, a transcriptional activator TenA homologue from hyperthermophilic archaeon *Pyrococcus horikoshii*. Acta Crystallogr Sect D 2004;60:1094–1100.
4. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
5. Terwillger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr Sect D 1999;55:849–861.
6. La Fortelle E, Bricogne G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol 1997; 276:472–494.
7. Terwillger TC. Maximum likelihood density modification. Acta Crystallogr Sect D 2000;56:965–972.
8. Terwillger TC. Automated main-chain model-building by template-matching and iterative fragment extension. Acta Crystallogr Sect D 2002;59:34–44.
9. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr Sect A 1991;47:110–119.
10. Zhou Y, Yao M, Tanaka I. New algorithm for protein model building: extending partial model in map segment. J Appl Crystallogr 2006;39:57–63.
11. Yao M, Zhou Y, Tanaka I. LAFIRE: software for automating the refinement process of protein-structure analysis. Acta Crystallogr Sect D 2006;62 (Part 2):189–196.
12. Brunger AT, Adams PD, Clore GM, Delano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, KusZzewski J, Nilges N, Pannun NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography and NMR system (CNS): a new software system for macromolecularstructure determination. Acta Crystallogr Sect D 1998;54:905–921.
13. Laskowski RA, Mac Arthur MW, Moss DS, Thornton JM. PROCHECK: a program to check stereochemical quality of protein structure. J Appl Crystallogr 1993;26:283–291.
14. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallograllogr 1991; 24:946–950.
15. Merritt EA, Bacon DJ. Raster3D: Photorealistic molecular graphics. Methods Enzymol 1997;277:505–524.
16. Holm L, Sander C. Dali: a network tool for proteins structure comparison. Trends Biochem Sci 1995;20:478–480.
17. Cook WJ, Kar SR, Taylor KB, Hall LM. Crystal structure of the cyanobacterial methallothionein repressor SmtB: a model for metalloregulatory proteins. J Mol Biol 1998;275:337–346.
18. Garcia-Castellanos R, Marrero A, Mallorqui-Fernandez G, Potempa J, Coll M, Gomis-Ruth FX. Three-dimensional structure of MecI. J Biol Chem 2003;278:39897–39905.
19. Alekshun MN, Levy SB, Mealy TR, Seation BA, Head J. The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution. Nat Struct Biol 2001;8:710–714.
20. Ramakrishnan V, Finch JT, Graziano V, Lee PL, Sweet RM. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature 1993;362:219–223.
21. Clark KL, Halay ED, Lai E, Burley SK. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature 1993;364:412–420.
22. Schwartz T, Roud MA, Lowenhaupt K, Herbert A, Rich A. Crystal structure of the Za domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. Science 1999;284:1841–1845.
23. Hietro A, Alexander SR, Kim J, Parg G, Emr SD, Hurley JH. Structure of the ESCRT-II endosomal trafiking complex. Nature 2004;431:221–225.
24. Safo M, Zhao Q, Ko T, Musayev FN, Robinson H, Scarsdale N, Wang AH, Archer GL. Crystal structure of the BlaI repressor from *Staphylococcus aureus* and its complex with DNA: insight into transcriptional regulation of the *bla* and *mec* operons. J Bacteriol 2005;187:1833–1844.
25. Zhang R, Kim Y, Skarina T, Beasley S, Laskowski R, Arrowsmith C, Edwads A, Joachimiak A, Savchenko A. Crystal structure of *Thermotoga maritima* 0065, a member of the IclR transcriptional factor family. J Biol Chem 2002;277:19183–19190.
26. Rice PA, Yang S, Mizuuchi K, Nash HA. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell 1996;87: 1295–1306.

 DOI 10.1002/prot