PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

Crystal structure of the conserved hypothetical protein TTHA1606 from *Thermus thermophilus* HB8

Fumiaki Tomoike,[1] Taisuke Wakamatsu,[1] Noriko Nakagawa,[2,3] Seiki Kuramitsu,[1,2,3] and Ryoji Masui[2,3]*

[1] Graduate School of Frontier Biological Sciences, Osaka University, Osaka 565-0871, Japan

[2] Department of Biological Sciences, Graduate School of Science, Osaka University, Osaka 560-0043, Japan

[3] RIKEN SPring-8 Center, Hyogo 679-5148, Japan

Key words: Nif3; conserved hypothetical protein; DNA-binding; YbgI; toroidal structure.

INTRODUCTION

NIF3 family proteins are highly conserved from bacteria to human [Fig. 1(A)], but their functions remain unclear. NIF3 was originally identified as a NGG1-interacting factor in yeast on two-hybrid screening.[1] Several studies of eukaryotic NIF3 homologues suggested that NIF3 family proteins take part in transcription.[2,3] The *ybgI* gene, the Nif3 family gene of *Escherichia coli*, is encoded in the operon including endonuclease VII, a DNA repair protein involved in oxidative base damages.[4] Microarray analyses indicated that the expression of *ybgI* was induced by treatment with a DNA damaging agent.[5] This suggests a possible function of YbgI in the DNA repair process.

The crystal structures of *E. coli* YbgI,[6] *Staphylococcus aureus* SA1388,[7] *Streptococcus pneumoniae* SP1609, and *Bacillus cereus* YqfO[8] have been determined. In these structures, two metal ions tightly bind in the cavity inside the hexameric troid. Although the coordination of the dinuclear center exhibits some similarity to those of other metal-binding proteins, it remains uncertain how the metal-binding site is involved in the function.

The *Thermus thermophilus* HB8 genome contains the gene encoding the NIF3 family protein TTHA1606, which consists of 242 amino acid residues. TTHA1606 is transcribed constantly[9] and translated into a protein under normal growth conditions (unpublished data). In the "Whole-Cell Project," we selected *T. thermophilus* HB8 for systematic study of the structures and functions of all proteins from a single organism.[10] In this research, we determined the structure of TTHA1606. In addition, we found that TTHA1606 has the ability to bind single-stranded DNA (ssDNA). Based on the structure, we discuss the DNA-binding site of TTHA1606.

METHODS

Protein overexpression and purification

The expression plasmid of THA1606 (RIKEN BioResource Center) was transformed into *E. coli* BL21(DE3). The transformants were cultured at 37°C to 1 × 10^8 cell/mL in 1.5 L of a medium containing LB broth supplemented with 50 μg/mL of ampicillin. The cells were then incubated for 5 h in the presence of isopropyl-β-D-thiogalactopyranoside and harvested by centrifugation. Cells (28 g) were lysed by sonication in 20 m*M* Tris-HCl (pH 8.0) and 50 m*M* NaCl, and the lysate was incubated at 70°C for 10 min. After centrifugation, TTHA1606 in

Grant sponsor: Ministry of Education, Science, Sports and Culture of Japan; Grant numbers: 17770089, 20570131.

*Correspondence to: Ryoji Masui, Department of Biological Sciences, Graduate School of Science, Osaka University, 1-1 Machikaneyama-cho, Toyonaka, Osaka 560-0043, Japan. E-mail: rmasui@bio.sci.osaka-u.ac.jp

Received 1 December 2008; Revised 19 December 2008; Accepted 17 January 2009
Published online 2 February 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22397

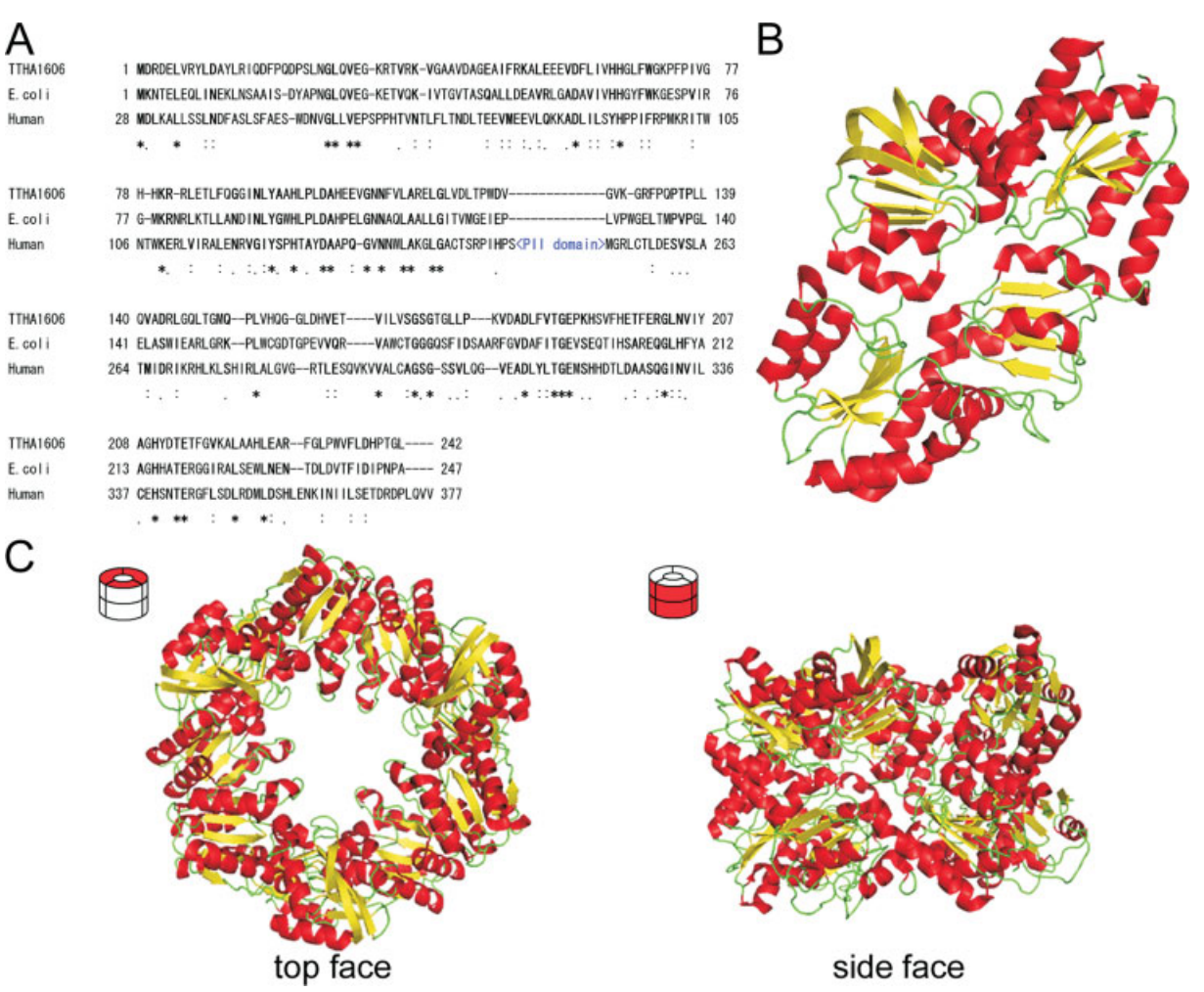

Figure 1

(**A**) Sequence alignment of TTHA1606, *E. coli* YbgI, and human NIF3LI. Human NIF3LI contains a PII domain in its middle region, which is omitted in this figure. Asterisks indicate conserved residues. (**B**) The structure of the dimer of TTHA1606 (viewed from the inside of the ring). (**C**) The structure of the hexamer of TTHA1606 viewed from two sides. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

the supernatant was purified by successive chromatographic steps with Resource ISO, Resource Q, BioScale CHT10-I hydroxyapatite, and HiLoad 16/60 200 pg columns. Purified proteins were stored in 20 m*M* Tris-HCl (pH 8.0).

Crystallization, data collection, and determination

The crystals of the TTHA1606 protein were obtained under condition number 14 (0.1*M* HEPES buffer, pH 7.5, containing 0.2*M* $CaCl_2$) of the Crystal Screen crystal screening kit (Hampton Research) using the 96-well sitting-drop vapor-diffusion method[11] at 20°C. A 0.5 µL aliquot of a crystallization reagent, consisting of 28% (w/w) PEG 400 and 0.1*M* HEPES buffer (pH 7.5), containing 0.2*M* $CaCl_2$, was mixed with 0.5 µL of a 25 mg/mL protein solution, and then the mixture was covered with 15 µL of silicone and paraffin oil. Data were collected with the RIKEN Structural Genomics Beamline II (BL26B2)[12] at SPring-8 (Hyogo, Japan). The data were processed using the HKL-2000 program suite.[13] The structure was solved using *E. coli* YbgI (PDB code 1NMP) as the model for molecular replacement using Molrep.[14] The automatic tracing procedure in the program ARP/wARP[15] was utilized to build the initial model. Model refinement was carried out using the programs Xtalview[16] and CNS.[17] According to PROCHECK,[18] the final model has 85.2% of the residues in the most favored region of the Ramachandran plot and no residues in the disallowed regions. Data-collection statistics and processed data statistics are presented in Table I. (The coordinates are available in the Protein Data Bank under accession code 2YYB and the nucleotide sequence

Table I
Data Collection and Refinement Statistics for TTHA1606

Crystal parameters	
Space group	*R*32
Unit-cell parameters (Å)	$a = b = 148.186$, $c = 196.395$
Data processing	
Resolution (Å)	50 to 2.60
No. of measured reflections	413,521
No. of unique reflections	25,740
Completeness (%)	100 (100)
Redundancy	16.1 (16.1)
$I/\sigma(I)$	53.3 (6.03)
R_{merge} (%)[a]	6.0 (35)
Refinement parameters	
Resolution range ($F > 0$) (Å)	50 to 2.60
No. of reflection	22,542
R[b]	0.2333
R_{free}	0.2801
No. of atoms	
Protein	3806
Water	20
Average B value (Å^2)	33.6
R.m.s.d.	
Bond lengths (Å)	2.0
Bond angles (°)	0.008

Values in parentheses are for the outermost shell.
[a] $R_{merge} = \Sigma_{hkl} \Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the average of individual measurements of $I_i(hkl)$.
[b] $R = ||F_0| - |F_c|| / \Sigma |F_0|$.

data reported are available in the DDBJ/EMBL/GenBank databases under the accession number DDBJ/EMBL/GenBank AP008226.)

Electrophoretic mobility-shift assay

A synthesized 37-mer oligonucleotide (5′-ATGTGAATC AGTATGGTTACTATCTGCTGAAGGAAAT-3′) was labeled at the 5′ end with [γ-^{32}P]ATP using T4 polynucleotide kinase. This oligonucleotide was used as ssDNA. Double-stranded DNA (dsDNA) was made by annealing the labeled oligonucleotide to the complementary oligonucleotide. The reaction was carried out at room temperature in 10 μL of 50 m*M* Tris-HCl and 100 m*M* KCl, pH 7.5, containing TTHA1606 and ssDNA or dsDNA. The mixture was electrophoresed and the bands in the gel were analyzed.

RESULTS AND DISCUSSION

The overall structure of TTHA1606 was determined at 2.6 Å resolution [Fig. 1(B)]. The asymmetric unit contains two TTHA1606 molecules. Each subunit consists of two interlinked α/β domains. Both domains comprise three layers, that is, two α-helix layers sandwiching one β-sheet. This fold is shared by all structurally characterized NIF3 family proteins.[6–8] The apparent molecular mass of the purified TTHA1606 was estimated to be ~180 kDa by gel filtration (data not shown), suggesting TTHA1606 forms a hexamer. Judging from the result of crystallographic analysis, it is possible for TTHA1606 to exist as a hexameric form, as shown in Figure 1(C). The overall shape of the hexamer can be viewed as a double-ring of two trimers: two trimers in face-to-face contact.

The results of electrophoretic mobility-shift assay (EMSA) showed that TTHA1606 bound to ssDNA but not dsDNA [Fig. 2(A)]. To our knowledge, this is the first report of the DNA-binding ability of a NIF3 family protein. Since the lysate of *E. coli* cells overexpressing TTHA1606 was heated at 70°C for 10 min during the purification process, it was unlikely that DNA-binding proteins of *E. coli* contaminated the purified preparations. It should be noted that the binding affinity varied from one preparation to another. Some factors in the preparations might affect the structural stability or activity of TTHA1606.

The troidal ring structure of TTHA1606 prompted us to examine the possibility that ssDNA binds to the inside of the ring. Several observations support this hypothesis. First, the inner surface of the TTHA1606 ring is more conserved as to amino acid sequence than the outside. We mapped levels of amino acid conservation to TTHA1606 structure with the ConSurf server.[19] The mapping of the dimer unit revealed that the inside of the ring was more conserved than the outside [Fig. 2(D)]. This suggests that the inside of the ring is important for the functional roles of NIF3 family proteins. In addition, the well-conserved residues are concentrated in and near the clefts found in the inner surface of the ring: each cleft is formed by two subunits [Fig. 2(B)]. Other structurally determined NIF3 proteins have metal ions in the cleft.[6–8] The conserved region at the interface of the dimer partially overlaps the metal-binding site in the cleft [Fig. 2(D)]. The purified TTHA1606 contained no metal ion, and the addition of metal ions had no apparent effect on the interaction with ssDNA (data not shown). Many DNA-binding proteins are known to bind DNA without metal ion. However, the metal-binding site of NIF3 family proteins is presumed to be important for unassigned functions. Metal ions may contribute to DNA sequence- or length-specific binding. Alternatively, metal ions may affect the binding to mRNA, modified nucleic acids, or polyphosphates.

Second, positively charged region exists near the metal-binding site in the cleft [Fig. 2(B)]. This region comprises another cavity on the opposite side of the cleft. This region is formed by residues of both subunits of the dimer [Fig. 2(C)]. The crystal structures of all NIF3 family proteins also have this cavity. It is possible that a negatively charged ligand binds to this cavity. Although many basic residues are found in this region [Fig. 2(C)], most of them are less conserved [Fig. 2(D)]. However, the side wall of the cleft near this cavity is composed of well-conserved residues [Fig. 2(D)]. Interestingly, Godsey *et al.* reported that a HEPES molecule bound to this cavity in the *B. cereus* YqfO structure and that this site is large enough to accommodate a nucleotide.[8] These observations imply that this cavity also comprises the ligand-binding site.

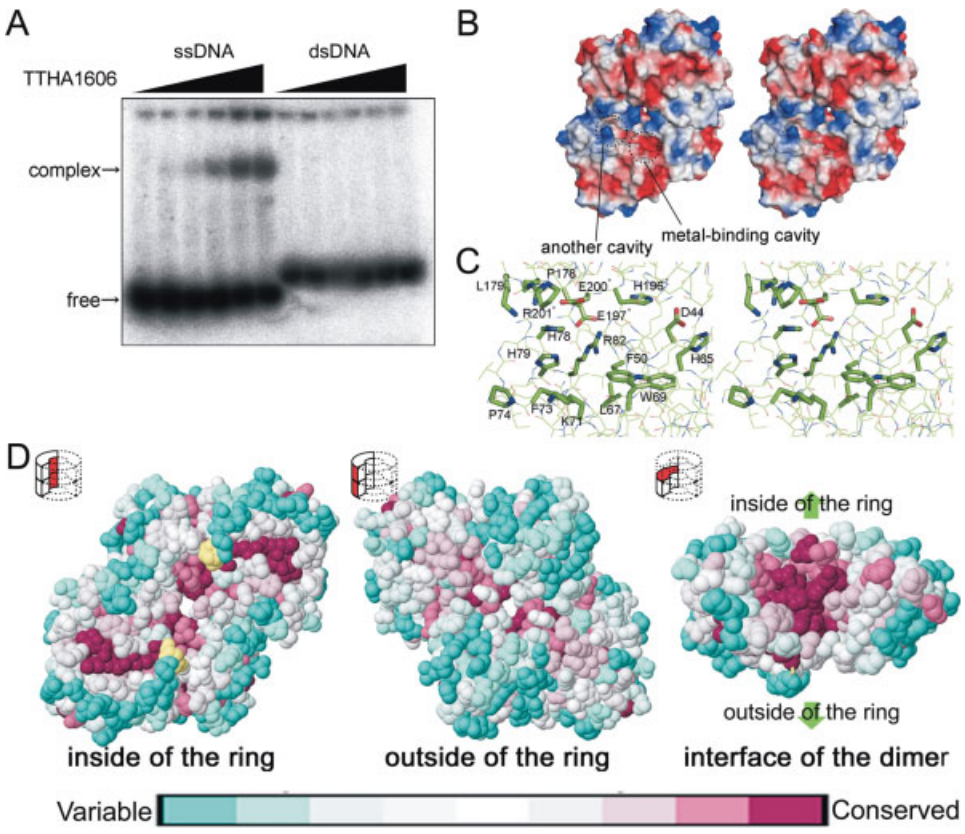

Figure 2

(**A**) The results of EMSA. The reaction mixture contained 10 n*M* ssDNA (the left six lanes) or dsDNA (the right six lanes) and TTHA1606 at 0, 4.5, 9, 18, 36, and 40 μ*M*. (**B**) Stereoview of the electrostatic potential map of TTHA1606 (viewed from the inside of the ring). Red and blue represent negatively and positively charged regions, respectively. Two cavities are also indicated. Compared with Figure 1(B) and the panel **D** in this figure, the dimer in this panel is viewed from a slightly different angle to bring the cleft into view. (**C**) Stereoview of another cavity indicated in panel B. Side chains are represented by sticks. Residue numbers followed by asterisks indicate residues of the other subunit in the dimer. (D) The conserved region of the TTHA1606 dimer. The amino acids are colored according to their conservation; see the color-coding bar.

Finally, TTHA1606 bound to ssDNA, but not to dsDNA. This might be due to restriction of the ligand size by the structure of the binding site. The inside diameter of the ring is ~24 Å, which is large enough for ssDNA to pass through it, but not for dsDNA.

B. cereus YqfO and *S. aureus* SA1388 contain an insert in the middle of the protein, which forms a PII-like domain in their crystal structures.[7,8] This domain forms trimeric lids that cover the openings on either side of the troidal rings. Godsey *et al.* suggested that some ligand might cause conformational change in these lids.[8] It should be also noted that the pattern of conserved residues are the same with or without the lids.

REFERENCES

1. Martens JA, Genereaux J, Saleh A, Brandl CJ. Transcriptional activation by yeast PDR1p is inhibited by its association with NGG1p/ADA3p. J Biol Chem 1996;271:15884–15890.
2. Tascou S, Uedelhoven J, Dixkens C, Nayernia K, Engel W, Burfeind P. Isolation and characterization of a novel human gene. NIF3L1, and its mouse ortholog Nif3l1, highly conserved from bacteria to mammals. Cytogenet Cell Genet 2000;90:330–336.
3. Akiyama H, Fujisawa N, Tashiro Y, Takanabe N, Sugiyama A, Tashiro F. The role of transcriptional corepressor Nif3l1 in early stage of neural differentiation via cooperation with Trip15/CSN2. J Biol Chem 2003;278:10752–10762.
4. Gifford CM, Wallace SS. The genes encoding endonuclease VIII and endonuclease III in *Escherichia coli* are transcribed as the terminal genes in operons. Nucleic Acids Res 2000;28:762–769.
5. Khil PP, Camerini-Otero RD. Over 1000 genes are involved in the DNA damage response of *Escherichia coli*. Mol Microbiol 2002;44:89–105.
6. Ladner JE, Obmolova G, Teplyakov A, Howard AJ, Khil PP, Camerini-Otero RD, Gilliland GL. Crystal structure of *Escherichia coli* protein ybgI, a toroidal structure with a dinuclear metal site. BMC Struct Biol 2003;3:7.
7. Saikatendu KS, Zhang X, Kinch L, Leybourne M, Grishin NV, Zhang H. Structure of a conserved hypothetical protein SA1388 from *S. aureus* reveals a capped hexameric toroid with two PII domain lids and a dinuclear metal center. BMC Struct Biol 2006;6:27.
8. Godsey MH, Minasov G, Shucaloca L, Brunzelle JS, Vorontsov II, Collart FR, Anderson WF. The 2.2 A resolution crystal structure of *Bacillus cereus* Nif3-family protein YqfO reveals a conserved dimetal-binding motif and a regulatory domain. Protein Sci 2007;16:1285–1293.

9. Shinkai A, Ohbayashi N, Terada T, Shirouzu M, Kuramitsu S, Yokoyama S. Identification of promoters recognized by RNA polymerase-σ^E holoenzyme from *Thermus thermophilus* HB8. J Bacteriol 2007;23:8758–8764.
10. Yokoyama S, Hirota H, Kigawa T, Yabuki T, Shirouzu M, Terada T, Ito Y, Matsuo Y, Kuroda Y, Nishimura Y, Kyogoku Y, Miki K, Masui R, Kuramitsu S. Structural genomics projects in Japan. Nat Struct Biol 2000;7:943–945.
11. Iino H, Naitow H, Nakamura Y, Nakagawa N, Agari Y, Kanagawa M, Ebihara A, Shikai A, Sugahara M, Miyano M, Kamiya N, Yokoyama S, Hirotsu K, Kuramitsu S. Crystallization screening test for the whole-cell project on *Thermus thermophilus* HB8. Acta Cryst 2008;F64:487–491.
12. Ueno G, Kanda H, Hirose R, Ida K, Kumasaka T, Yamamoto M. RIKEN structural genomics beamlines at the SPring-8; high throughput protein crystallography with automated beamline operation. J Struct Funct Genomics 2006;7:15–22.
13. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
14. Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. J Appl Cryst 1997;30:1022–1025.
15. Morris RJ, Perrakis A, Lamzin VS. ARP/wARP and automatic interpretation of protein electron density maps. Methods Enzymol 2003; 374:229–244.
16. McRee DE. A visual protein crystallographic software system for X11/XView. J Mol Graph 1992;10:44–46.
17. Brünger 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 and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.
18. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26:283–291.
19. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res 2005;33:W299–W302.