X-ray crystal structure of TTHB099, a CRP/FNR superfamily transcriptional regulator from Thermus thermophilus HB8, reveals a DNA-binding protein with no required allosteric effecto…

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · MAY 2012

Impact Factor: 2.63 · DOI: 10.1002/prot.24049 · Source: PubMed

3 AUTHORS:

Yoshihiro Agari
Illumina K.K. (Japan)
36 PUBLICATIONS **364** CITATIONS
SEE PROFILE

Seiki Kuramitsu
Osaka University
382 PUBLICATIONS **6,621** CITATIONS
SEE PROFILE

Akeo Shinkai
RIKEN
83 PUBLICATIONS **1,343** CITATIONS
SEE PROFILE

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

X-ray crystal structure of TTHB099, a CRP/FNR superfamily transcriptional regulator from *Thermus thermophilus* HB8, reveals a DNA-binding protein with no required allosteric effector molecule

Yoshihiro Agari,[1] Seiki Kuramitsu,[1,2] and Akeo Shinkai[1]*

[1] RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan

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

Key words: functional genomics; light-induced carotenoid biosynthesis; structural genomics; thermophile; transcription factor.

INTRODUCTION

Cyclic AMP (cAMP) receptor protein (CRP)/fumarate and nitrate reduction regulator (FNR) superfamily proteins are global transcriptional regulators that are widely distributed in bacteria and predominantly function as activators.[1,2] In many cases, CRP/FNR regulators respond to a wide range of endogenous and exogenous signals, such as cAMP, anoxia, redox state, oxidative and nitrosative stress, nitric oxide, carbon monoxide, 2-oxoglutarate, and temperature.[2] *Escherichia coli* CRP (also referred to as catabolite activator protein, CAP), a representative CRP family protein, is a dimer that undergoes a conformational change upon cAMP binding, and the CRP-cAMP complex interacts with DNA and RNA polymerase to regulate transcription.[3,4]

Thermus thermophilus HB8, which belongs to the phylum *Deinococcus-Thermus*, is an extremely thermophilic bacterium isolated from the water at a Japanese hot spring and grows at an optimum temperature range of 65°C–72°C.[5] Its genome is composed of the 1.85-Mbp chromosomal DNA, the 0.26-Mbp plasmid pTT27, and the 9.32-kbp plasmid pTT8, encoding 1,973, 251, and 14 open reading frames, respectively (NCBI accession numbers NC_006461, NC_006462, and NC_006463, respectively). Structural and functional genomics studies have been performed on this strain.[6] This strain has four CRP/FNR superfamily proteins: TTHA1437 (CRP), TTHA1567, TTHA1359 [stationary phase-dependent regulatory protein (SdrP)], and TTHB099 (NCBI accession numbers YP_144703, YP_144833, YP_144625, and YP_145338, respectively), which share 29%–39% amino acid sequence similarity with one another. *T. thermophilus* CRP positively regulates six promoters (22 genes), including clustered regularly interspaced short palindromic repeat-associated (Cas) genes,[7] in a cAMP-dependent manner. SdrP is an oxidative stress-responsive activator for 16 promoters (22 genes).[8,9] The activation of transcription by SdrP occurs independently of any added effector molecule *in vitro*, and this finding is supported by the observation that the three-dimensional structure of SdrP, with no effector molecules, is similar to that of the DNA-binding form of *E. coli* CRP.[8] Recently, Takano *et al.*[10] demonstrated the function of the TT_P0055 protein from *T. thermophilus* HB27, a homolog of the TTHB099 protein with one amino acid substitution. The *TT_P0055* gene is located in the *litR* operon on the megaplasmid pTT27, and its expression is regulated by the light-inducible transcriptional regulator, LitR.[10] Under light conditions, the expression of the

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Akeo Shinkai, RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan. E-mail: ashinkai@spring8.or.jp

Received 27 December 2011; Revised 18 January 2012; Accepted 25 January 2012
Published online 6 February 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24049

TT_P0055 gene is enhanced, and it positively regulates the expression of the adjacent *crtB* operon, which is involved in carotenoid biosynthesis.[10] Expression of TT_P0055 protein upregulates target genes *in vitro* without the need for an effector molecule or light.[10] In this study, we determined the X-ray crystal structure of the TTHB099 protein, which revealed that its three-dimensional structure is similar to those of the DNA-binding forms of CRP family proteins. The structure supports the observation that the TT_P0055 protein functions as a transcriptional activator, without the need for an effector molecule.

METHODS

Cloning, expression, and purification

The open reading frame of the *TTHB099* gene was cloned under the control of the T7 promoter (*Nde*I-*Bam*HI sites) in the *E. coli* expression vector pET-11a (Merck), to construct pET-tthb099 (RIKEN BioResource Center: http://www.brc.riken.jp/inf/en/index.shtml). The second codon of the open reading frame, AAG, was converted to AAA to optimize the codon usage in *E. coli*. The TTHB099 protein was produced in *E. coli* BL21 (DE3) cells (Merck) harboring pET-tthb099. The cell lysate was heated at 70°C for 13 min. The soluble fraction was then applied to a SuperQ Toyopearl 650 M column (Tosoh), pre-equilibrated with 20 m*M* Tris-HCl (pH 8.0). The flowthrough fractions were collected, and then applied to a HiTrap Heparin 5 mL column (GE Healthcare), pre-equilibrated with 20 m*M* Tris-HCl (pH 8.0), and the bound protein was eluted with a linear gradient of 0–2*M* NaCl. The target fractions were collected, and precipitated by 2.4*M* $(NH_4)_2SO_4$ containing 50 m*M* sodium phosphate (pH 7.0). The sample was dissolved in 20 m*M* MES buffer (pH 6.0) containing 1*M* NaCl, and then applied to a HiLoad 16/60 Superdex 75 pg column (GE Healthcare), pre-equilibrated with the same buffer. The target fractions were collected, concentrated to 1.2 mg/mL with a Vivaspin 20 concentrator (10,000 molecular-weight cutoff, Sartorius AG), and then buffer was exchanged to 20 m*M* MES buffer (pH 6.0), containing 0.6*M* NaCl and 1 m*M* dithiothreitol. The protein concentration was determined by measuring the absorbance at 280 nm.[11] To estimate the molecular mass of the purified TTHB099 protein, gel filtration chromatography was performed with a HiLoad 16/60 Superdex 75 pg column (GE Healthcare).

Crystallization, data collection, and structure determination

Crystallization of the TTHB099 protein was performed by the sitting drop vapor diffusion method, by mixing 1 μL of the protein solution with an equal volume of reservoir solution, followed by equilibration for 33 days against 0.5 mL of reservoir solution at 293 K. The best crystals were obtained with a reservoir solution containing 0.1*M* HEPES (pH 7.4), 18% polyacrylic acid 5100, and 0.1*M* $MgCl_2$. The diffraction data sets were collected at 1.92 Å resolution with the RIKEN Structural Genomics Beamline II (BL26B2)[12] at SPring-8 (Hyogo, Japan) (proposal number 20080039), using a MAR MX-225 CCD detector (Rayonix LLC). The collected data were processed with the HKL2000 program suite.[13] The molecular replacement was performed by the Molrep program (CCP4 Program Suite), using the *T. thermophilus* SdrP structure [Protein Data Bank (PDB) code: 2ZCW].[8] The model refinement, initial picking, and manual verification of water molecules were performed using the programs CNS and Coot.[14,15] The quality of the structure was validated using the ADIT! Validation Server in the PDBj (http://pdbdep.protein.osaka-u.ac.jp/validate/en/) and MolProbity (http://molprobity.biochem.duke.edu/). The atomic coordinates and structure factors have been deposited in the PDB, under the accession code 3B02. The data collection and refinement statistics are presented in Table I.

Table I
X-Ray Data Collection and Refinement Statistics

PDB ID	3B02
Data collection	
Wavelength (Å)	1.0
Resolution (Å)	50–1.92 (1.98–1.92)
Space group	$I4_132$
No. of molecules in an asymmetric unit	1
Unit cell parameters (Å, °)	$a = b = c = 147.63$ $\alpha = \beta = \gamma = 90$
No. of measured reflections	1,069,585
No. of unique reflections	21,153
Completeness (%)	99.5 (94.2)
Redundancy	50.6 (46.0)
I/σ (I)	88.3 (20.6)
R_{merge}[a] (%)	5.5 (24.8)
Refinement	
Resolution (Å)	46.7–1.92
R_{work}[b] (%)/R_{free}[c] (%)	18.3/22.0
No. of protein atoms/water atoms	1559/146
Wilson B factor (Å^2)	13.6
Average B-factors (Å^2) for protein/water	26.3/33.3
RMSD bond lengths (Å)	0.007
RMSD bond angles (°)	1.3
Ramachandran analysis[d]	
Favored (%)	97.4
Outliers (%)	0.00

Values in parentheses are for the highest-resolution shell.
[a] $R_{merge} = \Sigma_h\Sigma_i|I_{h,i} - \langle I_h\rangle|/\Sigma_h\Sigma_i I_{h,i}$, where $I_{h,i}$ is the ith measured diffraction intensity of reflection h and $\langle I_h\rangle$ is the mean intensity of reflection h.
[b] R_{work} is the R-factor $= \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.
[c] R_{free} is the R-factor calculated using 10% of the data that were excluded from the refinement.
[d] Calculated by MolProbity.

Other methods

The conserved domain and BLAST searches were performed on the http://www.ncbi.nlm.nih.gov/Structure/

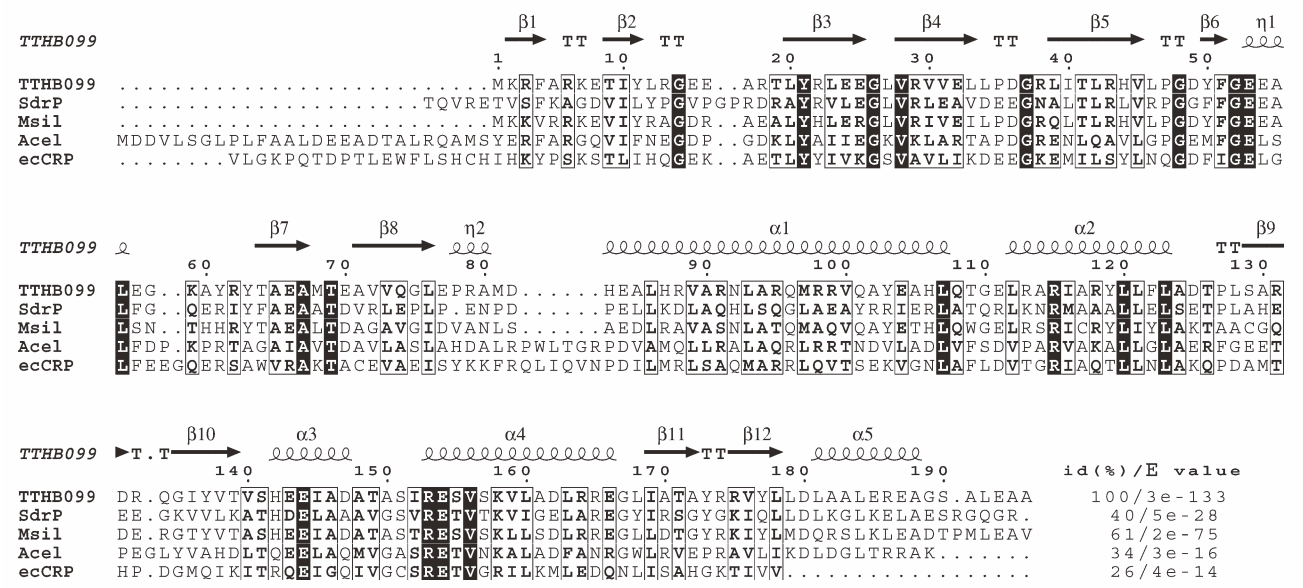

Figure 1

Sequence alignment of TTHB099 with representative homologous proteins. Strictly conserved residues are represented by white letters on a black background, and similar residues are depicted by boxed bold letters. SdrP, *T. thermophilus* SdrP[8]; Msil, *Meiothermus silvanus* DSM 9946 Mesil_1482 (YP_00368488.1); Acel, *Acidothermus cellulolyticus* 11B Acel_1995 (YP_873753.1); ecCRP, *E. coli* CRP.[1] The sequences were aligned using Clustal W2.[17] The secondary structure was predicted with DSSP,[18] and the figure was generated with ESPript 2.2.[19] α, η, β, and T represent α-helix, 3_{10}-helix, β-strand, and turn, respectively. The percentage identities [id(%)] and the E values relative to TTHB099, determined by BLAST, are indicated on the right.

cdd/wrpsb.cgi and http://blast.ncbi.nlm.nih.gov/Blast.cgi websites, respectively. The secondary-structure matching server is on the http://www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmserver website.

RESULTS AND DISCUSSION

The TTHB099 monomer is composed of 195 amino acid residues, with a calculated molecular mass of 22 kDa. Based on the results of the conserved domain search,[16] the protein has two conserved domains. One is an effector domain of the CAP family of transcription factors (cd00038), comprising residues K2–V100 with an E value of 3.03e−06 for the consensus sequence, and the other is a helix-turn-helix domain (cd00088), comprising residues R115–L179 with an E value of 5.18e−06 for the consensus sequence (Fig. 1). A BLAST search revealed that the homologous proteins that were most closely related to this protein, besides TT_P0055 from *T. thermophilus* HB27 (one amino acid substitution, E77D), were present in *T. thermophilus* SG0.5JP17-16 (2e−123), *T. aquaticus* Y51MC23 (2e−112), and *Meiothermus silvanus* DSM 9946 (2e−75). These homologs belong to the CRP/FNR superfamily (Fig. 1). *T. thermophilus* SdrP and *E. coli* CRP exhibited E values of 5e−28 and 4e−14, respectively (Fig. 1). The cysteine residues that exist in several CRP/FNR superfamily proteins that sense oxygen or redox variations[2] were not present in the TTHB099 protein (Fig. 1). The molecular mass of the purified protein, determined by gel filtration chromatography, was ∼45 kDa (Fig. S1), suggesting that it exists as a dimer in solution, as in the cases of the *E. coli* and *T. thermophilus* CRP proteins and SdrP.[1,7,8]

The three-dimensional structure of the TTHB099 protein was determined by the molecular replacement method, using the *T. thermophilus* SdrP structure (PDB code: 2ZCW),[8] and was refined to 1.92 Å resolution. The data collection and refinement statistics are summarized in Table I. The overall structure is shown in Figure 2(A). The N-terminal domain consists of two 3_{10}-helices (η1, residues 53–56; η2, residues 78–80) and eight β-strands (β1, residues 2–4; β2, residues 9–11; β3, residues 20–25; β4, residues 28–33; β5, residues 39–45; β6, residues 50 and 51; β7, residues 64–67; and β8, residues 71–76) that adopt a double-stranded β-helix fold with a jelly roll topology. The C-terminal domain consists of four α-helices (α2, residues 112–123; α3, residues 142–147; α4, residues 153–166; and α5, residues 181–188) and four β-strands (β9, residues 129–132; β10, residues 135–139; β11, residues 169–172; and β12, residues 175–178) that form a winged helix-turn-helix fold. These two domains are linked by a large α-helix (α1, residues 84–107).

The TTHB099 structure was compared with the previously determined structures in the PDB database, using the secondary-structure matching server.[21] The closest

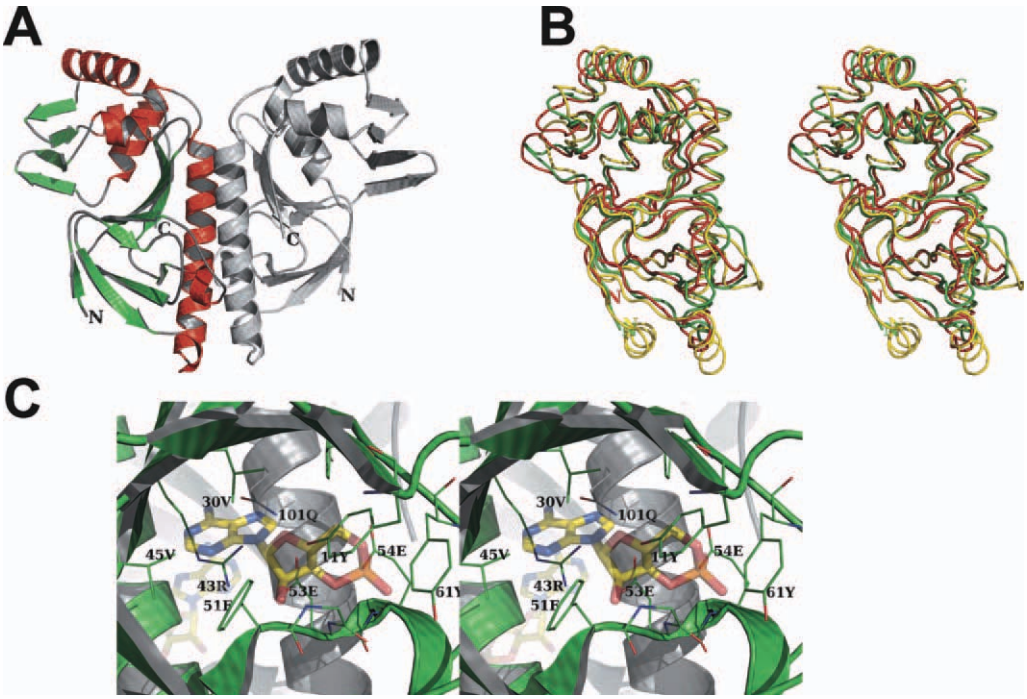

Figure 2

(**A**) Ribbon diagram of the TTHB099 dimer. The dimer molecule was generated around a crystallographic two-fold axis. The α-helices and β-strands in one chain are colored red and green, respectively, and the other chain is colored gray. (**B**) Stereoview of the superpositioned main chain structures of TTHB099 (red), SdrP (PDB code: 2ZCW) (green), and the DNA-binding form of *E. coli* CRP (PDB code: 2CGP chain A) (yellow). (**C**) Stereoview of the site within TTHB099 corresponding to the primary cAMP-binding site in *E. coli* CRP.[8,20] A cAMP molecule of the *E. coli* CRP-cAMP-DNA complex (PDB code: 2CGP chain A) is superimposed on TTHB099, as a transparent stick model. These figures were drawn using the PyMol program (http://www.pymol.org/).

structure was that of *T. thermophilus* SdrP,[8] with *Q*-, *P*-, and *Z*-scores of 0.68, 25.5, and 15.1, respectively, and a root mean square deviation (r.m.s.d.) of 1.70 Å [Fig. 2(B)]. SdrP is a transcriptional activator of CRP family proteins, and it does not require any effector molecule to bind DNA, unlike the case of *E. coli* CRP, which requires cAMP. The structure of the TTHB099 protein also resembled that of the DNA-binding form of *E. coli* CRP complexed with cAMP and DNA (PDB code: 2CGP chain A),[20] with *Q*-, *P*-, and *Z*-scores of 0.49, 10.2, and 10.2, respectively, and an r.m.s.d. value of 2.26 Å [Fig. 2(B)]. It should be noted that the side chains of the Y11, R43, E53, E54, and Q101 residues of the TTHB099 protein penetrate into a space corresponding to the cAMP-binding pocket of *E. coli* CRP [Fig. 2(C)]. No small molecules, except for water molecules, were found in the crystal structure of the TTHB099 protein. These results suggest that cAMP cannot enter the site corresponding to the primary cAMP-binding site of *E. coli* CRP without conformational change due to steric hindrance by the bulky residues, as in the case of SdrP, although the bulky residues differ from those of SdrP.[8]

The three-dimensional structure of the TTHB099 protein with no effector molecule resembled those of the DNA-binding forms of the CRP family transcriptional regulators, that is, SdrP and the *E. coli* CRP-cAMP complex. According to the structure, E77 is located between β8 and η2, and its side chain faces outside the molecule (Fig. S2), thus, the substitution of this residue with D, which is found in the TT_P0055 protein, may not affect the entire structure. Therefore, the structural properties of the TTHB099 protein imply that this protein does not require an effector molecule to bind DNA, and this proposal is supported by the observation that the TT_P0055 protein positively regulates transcription independently of an effector molecule *in vitro*.[10] Thus, the expression of the TT_P0055-regulated genes may depend on the amount of TT_P0055, as in the case of the expression of the SdrP-regulated genes, which depends on the amount of SdrP in *T. thermophilus* HB8.[9] The CRP family proteins may be categorized into the effector-dependent and -independent types, depending on the residues located within the space corresponding to the cAMP-binding pocket of *E. coli* CRP. In both the SdrP and TTHB099 proteins, several bulky residues penetrate into the spaces, although the residues differ between the two proteins.[8] Among the residues, Q101 of TTHB099 corresponds to Y107 of SdrP (Fig. 1). In the *E. coli* and *T. thermophilus*

CRP proteins, which require cAMP for activation, the corresponding residues, S129 and D131, respectively, are smaller.[8] Thus, the size of the residue at this position might be one of the critical features that determine the type of CRP.

T. thermophilus HB27 lacks the small plasmid pTT8, but the sequences of the chromosomal DNA and the megaplasmid pTT27 are quite similar to those of *T. thermophilus* HB8. *T. thermophilus* HB8 also has a light-responsive transcriptional regulator named CarH,[22] which is a homolog of *T. thermophilus* HB27 LitR, with one amino acid substitution (T218A). Furthermore, the gene organizations of the *litR* (*carH*) and *crtB* operons are almost the same between the two strains (NCBI accession numbers NC_006462 and NC_005838), and *T. thermophilus* HB8 is also yellow-pigmented.[23] Therefore, light-induced carotenoid production may also occur in *T. thermophilus* HB8, through the actions of the two transcriptional regulators, CarH and TTHB099.

ACKNOWLEDGMENTS

The authors thank Kayoko Matsumoto and Yoko Ukita for protein purification, Kazuko Agari and Aimi Osaki for crystallization, and Yuka Nonaka for data collection at SPring-8. They also thank Drs. Kenji Ueda and Hideaki Takano for critically reading the article.

REFERENCES

1. Kolb A, Busby S, Buc H, Garges S, Adhya S. Transcriptional regulation by cAMP and its receptor protein. Annu Rev Biochem 1993;62:749–795.
2. Körner H, Sofia HJ, Zumft WG. Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. FEMS Microbiol Rev 2003;27:559–592.
3. Sharma H, Yu S, Kong J, Wang J, Steitz TA. Structure of apo-CAP reveals that large conformational changes are necessary for DNA binding. Proc Natl Acad Sci USA 2009;106:16604–16609.
4. Popovych N, Tzeng SR, Tonelli M, Ebright RH, Kalodimos CG. Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. Proc Natl Acad Sci USA 2009;106:6927–6932.
5. Oshima T, Imahori K. Description of *Thermus thermophilus* (Yoshida and Oshima) com. nov., a non-sporulating thermophilic bacterium from a Japanese thermal spa. Int J Syst Bacteriol 1974;24:102–112.
6. 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(Suppl):943–945.
7. Shinkai A, Kira S, Nakagawa N, Kashihara A, Kuramitsu S, Yokoyama S. Transcription activation mediated by a cyclic AMP receptor protein from *Thermus thermophilus* HB8. J Bacteriol 2007;189:3891–3901.
8. Agari Y, Kashihara A, Yokoyama S, Kuramitsu S, Shinkai A. Global gene expression mediated by *Thermus thermophilus* SdrP, a CRP/FNR family transcriptional regulator. Mol Microbiol 2008;70:60–75.
9. Agari Y, Kuramitsu S, Shinkai A. Identification of novel genes regulated by the oxidative stress-responsive transcriptional activator SdrP in *Thermus thermophilus* HB8. FEMS Microbiol Lett 2010;313:127–134.
10. Takano H, Kondo M, Usui N, Usui T, Ohzeki H, Yamazaki R, Washioka M, Nakamura A, Hoshino T, Hakamata W, Beppu T, Ueda K. Involvement of CarA/LitR and CRP/FNR family transcriptional regulators in light-induced carotenoid production in *Thermus thermophilus*. J Bacteriol 2011;193:2451–2459.
11. Kuramitsu S, Hiromi K, Hayashi H, Morino Y, Kagamiyama H. Pre-steady-state kinetics of *Escherichia coli* aspartate aminotransferase catalyzed reactions and thermodynamic aspects of its substrate specificity. Biochemistry 1990;29:5469–5476.
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. 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 & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D 1998;54:905–921.
15. Emsley P, Cowtan K. *Coot*: model-building tools for molecular graphics. Acta Crystallogr D 2004;60:2126–2132.
16. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res 2002;30:281–283.
17. Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, Mcwilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics 2007;23:2947–2948.
18. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
19. Gouet P, Robert X, Courcelle E. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res 2003;31:3320–3323.
20. Passner JM, Steitz TA. The structure of a CAP-DNA complex having two cAMP molecules bound to each monomer. Proc Natl Acad Sci USA 1997;94:2843–2847.
21. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D 2004;60:2256–2268.
22. Ortiz-Guerrero JM, Polanco MC, Murillo FJ, Padmanabhan S, Elias-Arnanz M. Light-dependent gene regulation by a coenzyme B12-based photoreceptor. Proc Natl Acad Sci USA 2011;108:7565–7570.
23. Tian B, Hua Y. Carotenoid biosynthesis in extremophilic Deinococcus-Thermus bacteria. Trends Microbiol 2010;18:512–520.