Transcriptional regulatory mechanism of stress-responsive genes in *Thermus thermophilus* HB8

Yoshihiro Agari

Graduate School of Frontier Biosciences

Osaka University

Contents

Abbreviations

avr	average value
cAMP	cyclic adenosine monophosphate
Cas	CRISPR-associated
CRISPR(s)	clustered regularly interspaced short palindromic repeat(s)
CRP	cAMP receptor protein
EDTA	ethylenediaminetetraacetic acid
FNR	fumarate and nitrate reduction regulator
HTK	highly thermostable kanamycin nucleotidyltransferase (HTK) gene
IDE	sequence identity
MAD	multiple wavelength anomalous dispersion
MPD	2-methyl-2,4-pentanediol
ORF(s)	open reading frame(s)
PAGE	polyacrylamide gel electrophoresis
RAMP	repeat-associated mysterious protein
r.m.s.d.	root mean square deviation
RNAP	RNA polymerase
RT-PCR	reverse transcriptase-polymerase chain reaction
SeMet	selenomethionine

Se-SdrP	SeMet-containing SdrP
SdrP	stationary phase-dependent regulatory protein
SDS	sodium dodecyl sulfate

General Introduction

In all living organisms, from bacteria to higher eukaryotes, the environmental stress response system is essential for survival. For instance, the DNA repair system, one of the stress response systems in higher animals, may suppress carcinogenesis. In pathogenic bacteria, the stress response is involved in the expression of virulence and the acquisition of drug resistance for survival. Bacteria have fundamental and essential stress response mechanisms, because they are unicellular organisms and are close to the origin of life. Thus, studying bacterial stress response mechanisms will illuminate the fundamental systems of life.

Bacteria are exposed to various stresses in nature, including variations in nutrient availability, osmolality, redox, pH, and temperature, as well as antibiotics, toxic heavy metals, and invasion with foreign nucleic-acids, such as from bacteriophages, and so on. In order to adapt quickly to survive an abrupt environmental change, bacteria have developed an environmental response system to control the expression of specific proteins, to protect against various stresses and to repair damaged cellular components. Bacterial stress responses are mainly controlled at the transcriptional level (Browning & Busby, 2004; Gruber & Gross, 2003; Helmann, 2002; Hengge-Aronis, 2002; Hengge, 2008; Marles-Wright & Lewis, 2007; Raivio & Silhavy, 2001; Stock *et al.*, 2000). In bacteria, transcription is mediated by the sole RNA polymerase (RNAP) (Fig. 1). The prokaryotic RNAP holoenzyme is composed of the core enzyme, comprising the α, α, β, β' and ω subunits, and one of the several different σ factor species. Each σ factor is

responsible for promoter recognition and selectivity, while the core enzyme functions in the catalytic activity for RNA polymerization (Ebright, 2000; Sweetser *et al.*, 1987). Depending on the environmental conditions, an alternative σ factor, such as SigS, SigH or SigB, is activated, and as a consequence, a wide range of stress response genes are induced (Hecker & Volker, 2001; Marles-Wright & Lewis, 2007). In addition, stress response systems mediated by transcription factors have also been identified (Cheng Vollmer & Van Dyk, 2004). For instance, the two-component regulatory system, composed of the sensor histidine kinase and the response regulator, is one of the representatives of the bacterial stress response systems (Khorchid & Ikura, 2006; Kyriakidis & Tiligada, 2009; Stock *et al.*, 2000). The sensor histidine kinase recognizes a stress signal and becomes autophosphorylated, and then the histidine kinase

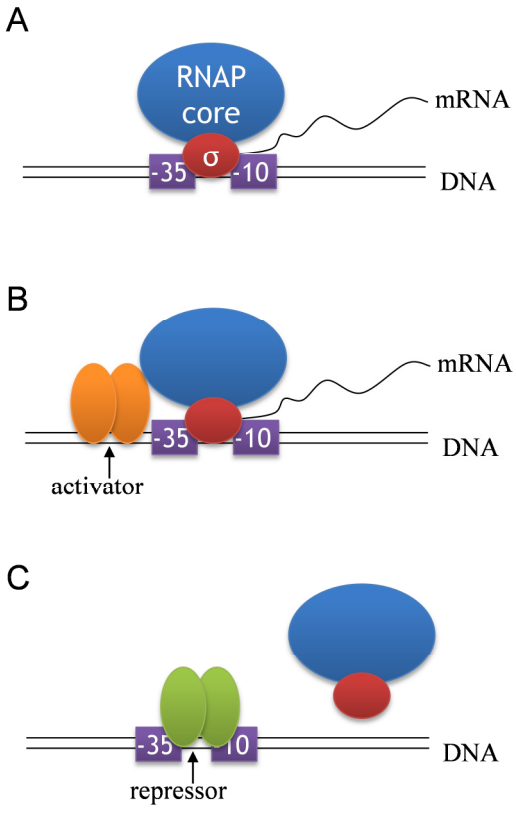


Fig. 1. Schematic diagrams of bacterial transcriptional regulation. **(A)** Regulator-independent transcription: RNA Polymerase (RNAP) synthesizes mRNA, based on the DNA sequence. **(B)** Activation: A transcriptional activator, which binds a specific DNA sequence in the upstream region of the promoter, interacts with RNAP to promote transcription. **(C)** Repression: A transcriptional repressor, which binds a specific DNA sequence on the promoter, inhibits RNAP binding to the promoter.

phosphorylates the cognate transcriptional regulator to control the transcription of the stress-responsive genes (Fig. 2A). *Escherichia coli* cAMP receptor protein (CRP) mediates catabolite repression in response to carbon source availability (Fig. 2B) (Epstein *et al.*, 1975). In many bacteria, the oxidative stress responsive genes are regulated by the LysR family transcriptional regulator OxyR (Schellhorn, 1995; Storz & Tartaglia, 1992) and the MerR family transcriptional regulator SoxR (Eiamphungporn *et al.*, 2006; Pomposiello & Demple, 2001). In *E. coli*, the GntR family transcriptional regulator FadR activates the *fab* genes, which are involved in unsaturated fatty acid biosynthesis in response to membrane damage (Rock & Cronan, 1996).

Thermus thermophilus HB8, which belongs to the phylum *Deinococcus-Thermus*, is a non-spore-forming, extremely thermophilic bacterium that was isolated from the water in a Japanese hot spring. The organism can grow at 47–85°C, with an optimum temperature range of 65–72°C, and its generation time under optimum conditions is 18–20 min (Oshima, 1974).

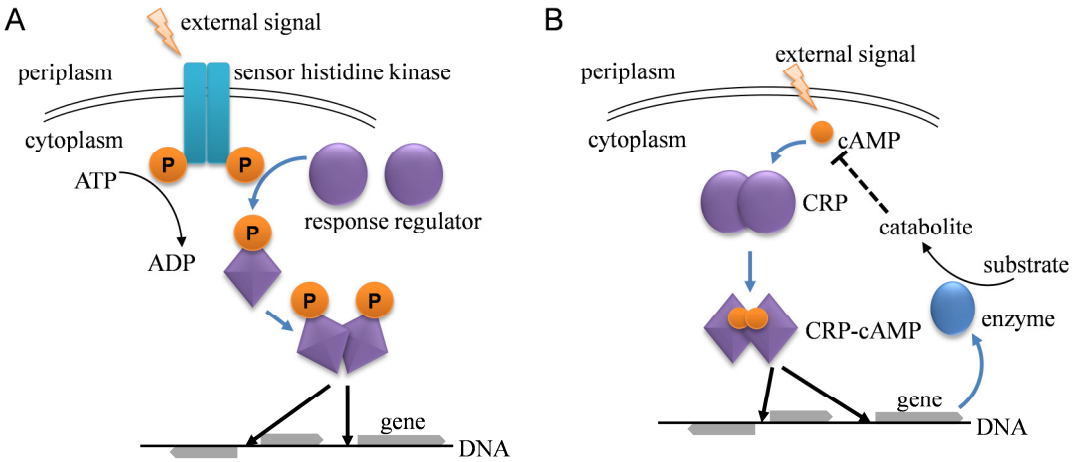


Fig. 2. Schematic diagrams of transcriptional regulation mediated by **(A)** a two-component system composed of a sensor histidine kinase and a response regulator, and **(B)** cAMP receptor protein (CRP).

The genome of this strain consists of the 1.85-megabase pair chromosomal DNA, the 0.26-megabase pair plasmid pTT27, and the 9.32-kilobase pair plasmid pTT8, which encode 1,973, 251, and 14 open reading frames (ORFs), respectively (GenBank Accession Nos. NC_006461, NC_006462, and NC_006463, respectively). Since the total number of genes (~2,200) is relatively small, as compared to those of *E. coli* (~5,000) and *Bacillus subtilis* (~4,000), which have been well studied as model organisms, and most proteins from *T. thermophilus* HB8 are highly stable for crystallization and functional studies, *T. thermophilus* HB8 is an ideal organism for structural and functional genomics studies (Yokoyama *et al.*, 2000a; Yokoyama *et al.*, 2000b). Furthermore, the number of transcriptional factors in *T. thermophilus* HB8 is relatively small (60–70), as compared with those in the cases of *E. coli* (~400) and *B. subtilis* (~300). Hence, the transcriptional regulatory network of *T. thermophilus* HB8 is relatively simple and fundamental, and this strain is a suitable model organism for studying the transcriptional regulatory mechanisms of stress-responsive genes. However, these mechanisms have barely been investigated so far.

I analyzed the two major stress responses of *T. thermophilus* HB8; i.e., that in the stationary growth phase and that against virus (phage) infection. In the course of my study, I also inspected the design of the *T. thermophilus* DNA microarray, which is used for genome-wide expression analyses of all mRNAs, and improved its accuracy for expression analysis.

Chapter 1

Stress response in the stationary growth phase

1-1 Abstract

Thermus thermophilus SdrP is one of four cyclic AMP receptor protein (CRP)/fumarate and nitrate reduction regulator (FNR) family proteins from the extremely thermophilic bacterium *T. thermophilus* HB8. The expression of *sdrP* mRNA increased in the stationary phase during cultivation at 70°C. Moreover, I found that *sdrP* mRNA expression was increased in response to various environmental or chemical stresses in the logarithmic growth phase, with oxidative stress being the most effective. Although the *sdrP* gene is non-essential, an *sdrP*-deficient strain showed growth defects, particularly when grown in synthetic medium, as well as increased sensitivity to disulfide stress. The expression of several genes was altered in the *sdrP* disruptant. Among them, I found eight SdrP-dependent promoters, using *in vitro* transcription assays. Furthermore, from a genome-wide expression pattern analysis using 306 DNA microarray datasets from 117 experimental conditions, eight additional SdrP-regulated genes were identified among the genes with expression that was highly correlated with that of *sdrP*. Based on the properties of the SdrP-regulated genes found in this study, these gene products may be involved in nutrient and energy supply, mRNA polyadenylation, DNA and protein repair and/or turnover, and redox control. The expression of the SdrP-regulated genes tended to increase upon entry into stationary phase. These results indicated that the main function of SdrP is in the oxidative stress response. A predicted SdrP binding site, similar to that recognized by *Escherichia coli* CRP, was found upstream of each SdrP-dependent promoter.

Transcriptional activation *in vitro* was independent of any added effector molecule. The hypothesis that apo-SdrP is the active form of the protein was supported by the observation that the three-dimensional structure of apo-SdrP is similar to that of the DNA-binding form of *E. coli* CRP (Agari *et al.*, 2008a; Agari *et al.*, 2010a). The expression levels of the SdrP-regulated genes probably depend on the concentration of SdrP, and not on post-translational modifications or effector binding to the protein.

1-2. Introduction

Bacterial cells enter the stationary phase upon nutrient depletion. During the stationary phase, the gene expression pattern changes globally, and the genes that are required for adaptation and survival, including those involved in nutrient scavenging, DNA repair, protein turnover and protection from oxidative damage, are expressed (Hengge-Aronis, 1996; Vicente *et al.*, 1999).

The CRP/fumarate and nitrate reduction regulator (FNR) superfamily proteins are global transcriptional regulators that are widely distributed in bacteria and predominantly function as activators (Kolb *et al.*, 1993; Korner *et al.*, 2003). 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 (Korner *et al.*, 2003). CRP and FNR are regarded as representative CRP/FNR family proteins. CRP undergoes a conformational change upon cAMP binding, and the CRP-cAMP complex interacts with DNA and RNAP to regulate transcription (Botsford & Harman, 1992; Busby & Ebright, 1999; Kolb *et al.*, 1993; Lawson *et al.*, 2004). Interestingly, CRP has diverse cellular roles in bacteria (Botsford & Harman, 1992). Crystallographic studies on *E. coli* CRP have been performed to determine the structures and the mechanisms underlying the interactions among CRP-cAMP, DNA and the C-terminal domain of the RNAP α subunit (Lawson *et al.*, 2004). The crystal structure of *E. coli* CRP revealed that two cAMP

binding sites are present in each monomer (Passner & Steitz, 1997). One is the primary binding site located in the N-terminal domain, which exhibits micromolar affinity for cAMP, and cAMP binding to this site leads to an allosteric change in CRP that allows DNA binding (Fig. 3) (Passner *et al.*, 2000; Scott & Jarjous, 2005; Sharma *et al.*, 2009). The other is the secondary binding site, lying between the N- and C-terminal domains, which has millimolar affinity for cAMP (Heyduk & Lee, 1989). DNA binding to CRP depends on the cAMP concentration (Heyduk & Lee, 1989; Scott & Jarjous, 2005). FNR senses oxygen via an iron-sulfur cluster ligated through cysteine residues (Green *et al.*, 2001; Kiley & Beinert, 2003). Hence, under anaerobic conditions, FNR is able to bind to specific DNA targets at promoters and thereby regulate transcription. In contrast, under aerobic conditions, FNR is converted to a form that is unable to bind to these targets. In *E. coli,* YeiL has been identified as a CRP/FNR-family transcription factor that is expressed in the stationary phase. This protein is believed to function as a nitrogen starvation regulator, because it may have an iron-sulfur center and reversible intra- and interchain disulfide bonds, as seen in FNR family proteins (Anjum *et al.*, 2000).

In *T. thermophilus* HB8, four ORFs; i.e., TTHA1437, TTHA1567, TTHA1359, and TTHB099 (NCBI Accession Nos. YP_144703, YP_144833, YP_144625, and YP_145338, respectively), which share 29–39% amino acid sequence similarity with one another, have been identified as CRP/FNR family proteins. I found that the expression of one of the CRP/FNR

family proteins, TTHA1359 (subsequently named SdrP: Stationary phase-dependent regulatory protein), increased upon entry into the stationary phase, and that this protein regulated many stress-responsive genes (Agari *et al.*, 2008a; Agari *et al.*, 2010a). In this chapter, I will describe the transcriptional regulatory mechanism via SdrP.

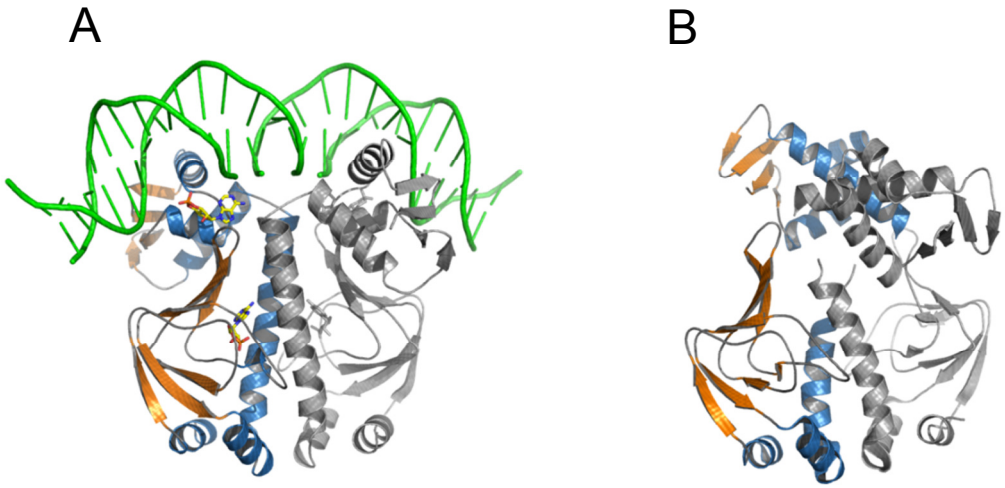


Fig. 3. Ribbon diagrams of the active and inactive forms of *E. coli* CRP. The α-helices and β-strands in one chain are colored blue and orange, respectively, the other chain is colored gray, and DNA in active form is colored green. **(A)** Active form: CRP-cAMP complex binding with DNA (Passner & Steitz, 1997). **(B)** Inactive form: apo-CRP (Sharma *et al.*, 2009).

1-3. Materials and methods

Overproduction and purification of recombinant SdrP

The *T. thermophilus sdrP* (*TTHA1359*) gene was amplified by genomic PCR using primers P01 and P02 (Table 1), and the amplified fragment was cloned under the control of the T7 promoter (NdeI–BamHI sites) of *E. coli* expression vector pET-11a (Merck, Darmstadt, Germany) to construct pET11a-SdrP. *E. coli* BL21(DE3) (Merck) harboring pET11a-SdrP was cultured at 37°C in 6 l of LB broth containing ampicillin (50 mg ml^{-1}) for 16 h. The cells were re-suspended in 70 ml of buffer containing 20 mM Tris-HCl (pH 8.0), 50 mM NaCl and 25 mM 2-mercaptoethanol and disrupted by sonication in ice water. The same volume of buffer pre-heated at 70°C was added to the cell lysate; this was followed by incubation for 10 min at 70°C and then ultracentrifugation (200,000 g) for 1 h at 4°C. Ammonium sulfate was added to the supernatant to a final concentration of 1.5 M, and the solution was applied to a RESOURCE PHE column (GE Healthcare UK, Buckinghamshire, UK) pre-equilibrated with 50 mM sodium phosphate buffer (pH 7.0) containing 1.5 M ammonium sulfate. The flow-through fractions were collected, and ammonium sulfate was added to a final concentration of 2.4 M. The precipitate was collected by centrifugation, suspended in 10 ml of 20 mM Tris-HCl (pH 8.0) and desalted by fractionation on a HiPrep 26/10 desalting column (GE Healthcare UK). The sample was then applied to a RESOURCE Q column (GE Healthcare UK) pre-equilibrated with 20 mM Tris-HCl (pH 8.0). The bound protein was eluted with a linear gradient of 0–1 M NaCl

in 20 mM Tris-HCl (pH 8.0). The target fractions were collected, desalted, and applied to a hydroxyapatite column BioScale CHT10-I (Bio-Rad Laboratories, Hercules, CA, USA) that had been pre-equilibrated with 10 mM sodium phosphate buffer (pH 7.0) containing 150 mM NaCl. The bound protein was eluted with a linear gradient of 10–250 mM sodium phosphate buffer (pH 7.0) containing 150 mM NaCl. The target fractions were collected, concentrated with a Vivaspin 20 concentrator (10,000 molecular-weight cut-off; Sartorius AG, Goettingen, Germany), and applied to a HiLoad 16/60 Superdex 75 pg (GE Healthcare UK) column that had been pre-equilibrated with 20 mM Tris-HCl (pH 8.0) containing 150 mM NaCl. The target fractions were collected and concentrated with a Vivaspin 20 concentrator (10,000 molecular-weight cut-off; Sartorius AG). The protein concentration was determined by measuring the absorbance at 280 nm (Kuramitsu & Yoshida, 1990).

Selenomethionine (SeMet)-containing SdrP (Se-SdrP) was generated using the methionine auxotroph *E. coli* Rosetta834(DE3), which was obtained by introducing the pRARE plasmid (Merck) into the B834(DE3) strain (Merck) as the host. The recombinant strain was grown in LeMaster medium (LeMaster & Richards, 1985) containing 50 mg ml^{-1} SeMet, 1.0% lactose, 50 mg ml^{-1} ampicillin and 30 mg ml^{-1} chloramphenicol. Se-SdrP was purified by the same process as the native protein (see above).

Table 1. Oligonucleotides used in this study

Oligonucleotide	Sequence (5′→3′)*	Used for
P01 P02	fwd, ATATcatatgACCCAGGTGCGCGAGACCGTGAGCTTCA; rev, ATATggatccTTATTAGCGCCCCTGGCCCCGGCTTTCGGCG	cloning of *sdrP* gene
P03 P04	fwd, ATATAggatccGGGGAGCGCCCGGGAGGCCTTGGGGCTTA; rev, ATATgaattcGCATGCCCCCAATATACGCCTGAGCCTGTA	cloning of upstream region of the *TTHA0337* gene
P05 P06	fwd, ATATAggatccTCCGCTGAGCTAGCCGCCCGCAACCGCTA; rev, ATATgaattcCGGGGGCACTATACCCTTGAGCCTGTGGAT	cloning of upstream region of the *TTHA1028* gene
P07 P08	fwd, ATATAggatccCGCCGCTTGCGCCACTCGGGGTACACGAGA; rev, ATATgaattcATACCGGGCCTAGAGTATACCCCAGGATAG	cloning of upstream region of the *TTHA0654* gene
P09 P10	fwd, ATATAggatccCGGACGGGAGGGGCGCAAGGGCAGGGGGCT; rev, ATATgaattcATACGCCGGGGAGTCTAACCCGGGGGGAGG	cloning of upstream region of the *TTHA0425* gene
P11 P12	fwd, ATATAggatccGGCCTGTTGCGCGGCGAAGCAGGCCCAGGC; rev, ATATgaattcCTTCACCCTTACCAGTCTAGGGGAAGGGAT	cloning of upstream region of the *TTHA0634* gene
P13 P14	fwd, ATATAggatccGGCCTTCGGCCTTCTGGCCTCGGG; rev, ATATgaattcACCAGCCACCCCCTTCCCCGACAGG	cloning of upstream region of the *TTHA0570* gene
P15 P16	fwd, ATATAggatccGCCAAAGCGCCTCCCCGACCACGCCCCCCATTGTAAGGAGGAGGCTTT; rev, ATATgaattcTCCCCTCCAGGCTAGGGCGGGAGGCCGAAGGGGAAAAGCCTCCTCCTTA	cloning of upstream region of the *TTHA0986* gene
P17 P18	fwd, ATATAggatccCTGGCGGCCGCGGTCCCCGGCCAAAGCCTGAACGCCCATTGTGCCCCGGGGTGC; rev, ATATgaattcTCTGCCCCTATGATACCCCCCTTTCAGGAAACAAAGGCACCCCGGGGCACA	cloning of upstream region of the *TTHA0770* gene
P19; P20	fwd, TGCATGCCTGCAGGTCGACT; rev, GATCGGTGCGGGCCTCTTCG	preparation of template DNA for the transcription assay
P21; P22	fwd, GATCGGTGCGGGCCTCTTCG; rev, CGCCGCGGTGGGGGCCCGGG	preparation of *TTHA0337* template a
P23; P24	fwd, GATCGGTGCGGGCCTCTTCG; rev, GGGGGCCCGGGGATTGTGCC	preparation of *TTHA0337* template b
P25; P26	fwd, GATCGGTGCGGGCCTCTTCG; rev, CCGGGGATTGTGCCTTTTAC	preparation of *TTHA0337* template c
P27; P28	fwd, GATCGGTGCGGGCCTCTTCG; rev, GATTGTGCCTTTTACCCCTG	preparation of *TTHA0337* template d
P29; P30	fwd, GATCGGTGCGGGCCTCTTCG; rev, CCTTTTACCCCTGACACATA	preparation of *TTHA0337* template e
P31; P32	fwd, GATCGGTGCGGGCCTCTTCG; rev, GGTTGTGCCTCGCCTGCTCC	preparation of *TTHA0634* template a
P33; P34	fwd, GATCGGTGCGGGCCTCTTCG; rev, CCTCGCCTGCTCCAGCCAGT	preparation of *TTHA0634* template b
P35; P36	fwd, GATCGGTGCGGGCCTCTTCG; rev, TCCCACGCTCCAGGGGAACC	preparation of *TTHA0634* template c
P37; P38	fwd, GATCGGTGCGGGCCTCTTCG; rev, CGCTCCAGGGGAACCCACGG	preparation of *TTHA0634* template d
P39; P40	fwd, GATCGGTGCGGGCCTCTTCG; rev, GGGGAACCCACGGCACACAC	preparation of *TTHA0634* template e

Continued on next page.

Table 1. Continued

Oligonucleotide	Sequence (5′→3′)*	Used for
P41; P42	fwd, CAGGTGCGCGAGACCGTGAGCTTCA; rev, CAGCTCGTCGTGGGTGGCCTTGAGC	RT-PCR analysis
P43; P44	fwd, ATAggatccCAGCCGCGCCGTGATCGGGCA; rev, TATgaattcCTCGCCCATAGGGGAAGCCTA	cloning of upstream region of the *TTHA0029* gene
P45; P46	fwd, ATAggatccCAAGCCCCACGAGGGGGCGTA; rev, TATgaattcGTACGGCATACTTCACCTCCG	cloning of upstream region of the *TTHA0557* gene
P47; P48	fwd, ATAggatccGTCTACCACGCCATCGGCCTG; rev, TATgaattcGAGGCGCATGCCCCCAGGCTA	cloning of upstream region of the *TTHA1128* gene
P49; P50	fwd, ATAggatccGCTTCAGGCCAGGAGGTCCGA; rev, TATgaattcGTCCACCATGGCCCTGAGTTT	cloning of upstream region of the *TTHA1215* gene
P51; P52	fwd, ATAggatccGCGCCGTGGGAGGCGGGGGCT; rev, TATgaattcGACCGGCATAGAACCTCCCTT	cloning of upstream region of the *TTHA1625* gene
P53; P54	fwd, ATAggatccCCTGCACCGCCCGTAGCTCCA; rev, TATgaattcTTCCACCATGCCGAACCTCCT	cloning of upstream region of the *TTHA1635* gene
P55; P56	fwd, ATAggatccCTAACCCTGGCGGTCCCGCCC; rev, TATgaattcGAAGGTCATCCTCACCCAGTA	cloning of upstream region of the *TTHA1892* gene
P57; P58	fwd, ATAggatccCCATCTTCTTGCCCTCCTTTC; rev, TATgaattcTTCAGGCATAGCCCCAGGTTA	cloning of upstream region of the *TTHB132* gene
P59; P60	fwd, ATAggatccTTACACCCGCTTGTGCGTTCT; rev, TATgaattcTTCGGGCATGGGGTCCTCCTT	cloning of upstream region of the *TTHA0987* gene

*Restriction sites are in lower case. fwd, forward; rev, reverse.

Crystallization

The crystallization conditions for Se-SdrP in the hanging-drop vapor diffusion method were determined using Crystal Screen Kits (Hampton Research, Aliso Viejo, CA, USA) at 293 K. Two microliters of Se-SdrP [11.6 mg ml^{-1} in 20 mM Tris-HCl (pH 8.0) containing 1 mM dithiothreitol] was mixed with the same volume of the reservoir solution, followed by equilibration against 0.5 ml of the reservoir solution. The best conditions were obtained with a reservoir solution containing 31% (v/v) 2-methyl-2,4-pentanediol (MPD), 0.1 M sodium acetate trihydrate (pH 5.1) and 20 mM calcium chloride dihydrate.

X-ray diffraction data collection and structure determination

A crystal of Se-SdrP was mounted on a cryoloop and flashcooled in a nitrogen gas stream at 100 K. The multiple wavelength anomalous dispersion (MAD) data were collected at three different wavelengths with an R-AXIS V detector (Rigaku Americas, Woodlands, TX, USA) using Structural Genomics Beamline II (BL26B2) at SPring-8 (Hyogo, Japan). The oscillation angle was 1°, exposure time was 10 s per frame and camera distance was 300 mm. All diffraction images were processed using the HKL2000 program suite (Otwinowski & Minor, 1997).

Selenium sites were determined with the SOLVE program (Terwilliger & Berendzen, 1999), and the resulting phases were improved with the RESOLVE program (Terwilliger & Berendzen,

1999). The initial model was built with the ARP/wARP program (Perrakis *et al.*, 2001), and further manual model building was performed using XtalView/Xfit (McRee, 1999). Simulated annealing, energy minimization and B factor refinement were carried out using the CNS program package (Brünger *et al.*, 1998). Cycles of manual modeling and CNS refinement were performed; 10% of the total reflections were randomly chosen for the R_{free} sets. The quality of the structure was analyzed using PROCHECK in the CCP4 suite (Collaborative Computational Project, 1994).

Disruption of the *T. thermophilus sdrP* gene

The *sdrP*-disruptant (Δ*sdrP*) strain was isolated by basically following the same process as that described previously (Hashimoto *et al.*, 2001). I constructed plasmid pGEM-Δ*sdrP*, which carries the upstream region (positions 1,297,481–1,298,001 of chromosomal DNA) of the *sdrP* gene followed by the thermostable kanamycin-resistance marker (*HTK*) gene and the downstream region (positions 1,298,567–1,299,086 of chromosomal DNA) of the *sdrP* gene, as described below (Fig. 4A). The fragments containing the upstream (fragment A) and downstream (fragment B) regions were each amplified by genomic PCR. The 3′-terminal 17 bp sequence of fragment A was derived from the 5′-terminal of the *HTK* gene, and the 5′-terminal 19 bp sequence of fragment B was derived from the 3′-terminal of the *HTK* gene. Fusion PCR was performed using fragments A and B and a plasmid carrying the *HTK* gene

(Hashimoto *et al.*, 2001). The amplified fragment was ligated with the pGEM-T Easy vector (Promega, Madison, WI, USA) to construct pGEM-Δ*sdrP*. pGEM-Δ*sdrP* was transformed into the *T. thermophilus* HB8 strain, and a kanamycin-resistant clone was isolated as the Δ*sdrP* strain (Fig. 4B). Genomic PCR and Southern hybridization were used to confirm that the *sdrP* gene had been replaced with the *HTK* gene.

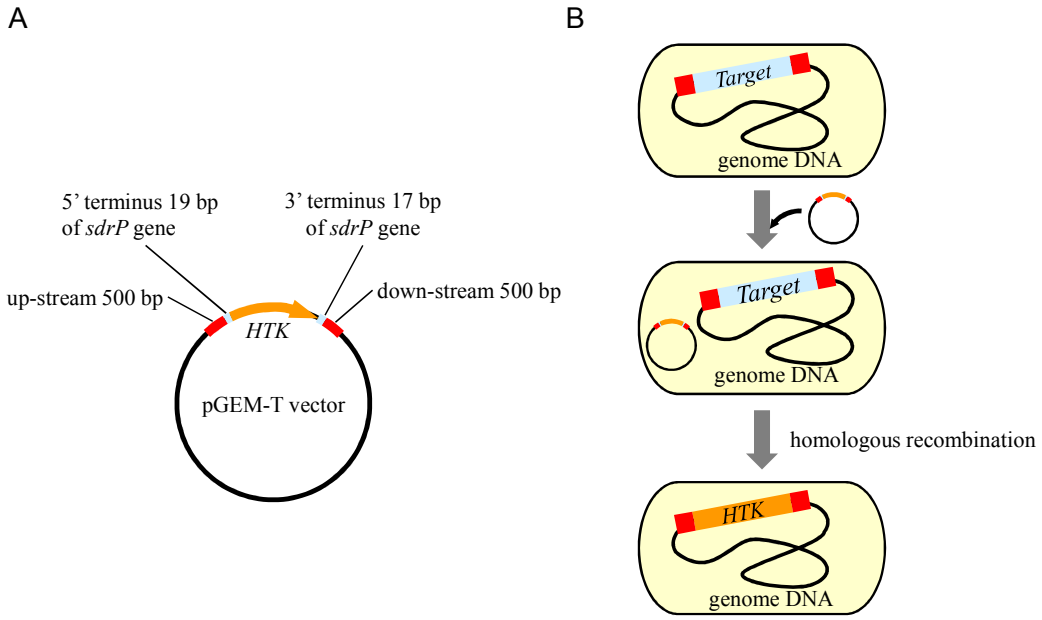


Fig. 4. (A) Schematic diagram of the plasmid vector for gene disruption. **(B)** Schematic diagram of gene disruption by homologous recombination.

Media and growth conditions for *T. thermophilus* HB8

TT broth containing 0.8% polypeptone, 0.4% yeast extract, 0.2% NaCl, 0.4 mM $CaCl_2$, and 0.4 mM $MgCl_2$ was adjusted to pH 7.2 with NaOH and used as the rich medium. Synthetic medium was prepared by mixing 500 ml of solutions A (4% sucrose) and B, 5 ml of solution C (2.5% $MgCl_2 \cdot 6H_2O$ and 0.5% $CaCl_2 \cdot 2H_2O$), 1 ml of solution D (1% biotin and 10% thiamine),

and 0.1 ml of each metal solution. Solution B was composed of 4% sodium glutamate, 0.11% K_2HPO_4, 0.036% KH_2PO_4, 0.4% NaCl, and 0.1% $(NH_4)_2SO_4$ and was adjusted to pH 7.3 with 3 M NaOH. The metal solutions used were as follows: 10% $FeSO_4 \cdot 7H_2O$, 1.2% $Na_2MoO_4 \cdot 2H_2O$, 0.1% $VOSO_4 \cdot xH_2O$, 0.5% $MnCl_2 \cdot 4H_2O$, 0.06% $ZnSO_4 \cdot 7H_2O$, 0.015% $CuSO_4 \cdot 5H_2O$, 0.8% $CoCl_2 \cdot 6H_2O$, and 0.02% $NiCl_2 \cdot 6H_2O$; all of these were sterilized by filtration.

For cultivation in the TT broth, the *T. thermophilus* HB8 strain was pre-cultured at 70°C for 16 h in 3 ml of broth. Two milliliters of culture broth was inoculated into 1 l of TT broth followed by cultivation at 70°C. For cultivation in the synthetic medium, the strain was pre-cultured at 70°C for 16 h in 3 ml of synthetic medium, and 2 ml of culture medium was then inoculated into 250 ml of synthetic medium, followed by cultivation at 70°C.

For the reverse transcriptase (RT)-PCR and expression pattern analyses, various physical or chemical stresses, such as *N*,*N*,*N*′,*N*′-tetramethylazodicarboxamide [diamide, (Tokyo Chemical Industry, Tokyo, Japan)], which leads to the formation of non-native disulfide bonds in proteins (Fig. 5) (Leichert *et al.*, 2003; Nakunst *et al.*, 2007), as well as H_2O_2, $ZnSO_4$, $CuSO_4$, $FeSO_4$,

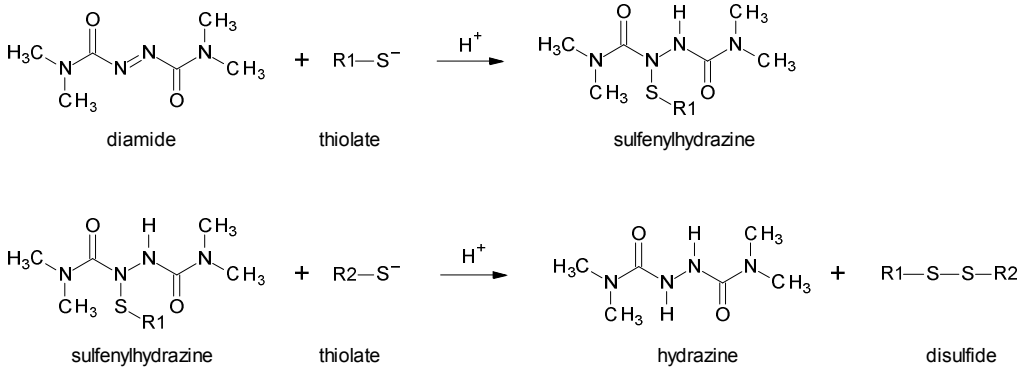


Fig. 5. Schematic diagram of the diamide reaction with free thiols (Leichert *et al.*, 2003).

tetracycline, NaCl, ethanol, and heat (80°C) stresses, were analyzed (see "RT-PCR analysis" subsection).

DNA microarray analysis

I used the TTHB8401a520105F GeneChip, which has oligonucleotide probes corresponding to the entire genomic sequence of *T. thermophilus* HB8. The basal probe design and the detection system of the GeneChip are described in the "Improvement of the Chip Definition File of the *Thermus thermophilus* HB8 GeneChip" subsection of chapter 2.

(i) *Sample preparation and data collection.* Sample preparation and data collection for the DNA microarray analysis were performed by basically the same procedure as that described previously (Agari *et al.*, 2008a; Shinkai *et al.*, 2007a) (Fig. 6). The procedure is as follows. The *T. thermophilus* HB8 strain was cultured in TT broth or synthetic medium at 70°C. Cells were collected from 5 to 200 ml of the culture medium, and then the crude RNA was extracted by the addition of 1.4 ml of a solution comprising 5 mM Tris-HCl, pH 7.5, 5 mM EDTA, 0.25% sodium dodecyl sulfate (SDS), and 50% of water-saturated phenol. This mixture was incubated at 65°C for 5 min, chilled on ice for 5 min, and then centrifuged at 4°C. A 750 µl aliquot of TRIZOL LS (Invitrogen, Carlsbad, CA, USA) was then added to 0.2 ml of the aqueous phase. After an incubation for 5 min at room temperature, the RNA was extracted with 0.2 ml of chloroform. The extraction was repeated with 0.5 ml of chloroform, and the

aqueous phase was precipitated with isopropanol. The pellet was dissolved in 0.2 ml of nuclease-free water, precipitated with ethanol, and then resuspended in 0.2 ml of water. The RNA was treated with DNase I (Invitrogen) at 37°C for 20 min in a 25 μl reaction mixture. The reaction was terminated by the addition of 1 μl of 0.5 M EDTA, followed by an incubation at 70°C for 5 min. The cDNA was synthesized with SuperScript II (Invitrogen) reverse transcriptase, in the presence of the RNase inhibitor SUPERase·In (Invitrogen) and 6-base random primers (Invitrogen). The cDNA was fragmented with 35 units of DNase I (GE Healthcare UK) at 37°C for 10 min, and after inactivation at 98°C for 10 min, the cDNA fragments were labeled with biotin-dideoxy UTP, using DNA labeling reagents from Affymetrix (Affymetrix, Santa Clara, CA, USA) or ENZO (ENZO Life Sciences, Farmingdale, NY, USA) and terminal transferase, according to the manufacturer's instructions.

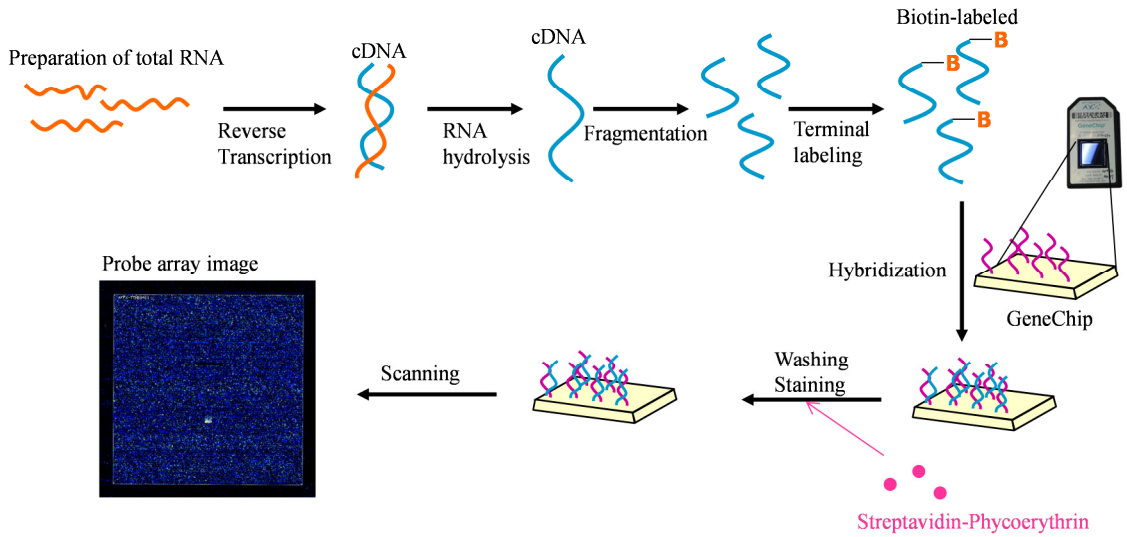


Fig. 6. Experimental procedure for the DNA microarray analysis.

The 3′-terminally labeled cDNA (2 μg) was hybridized to the GeneChip (Affymetrix). The

array was incubated for 16 h at 50°C, in a solution comprising 180 mM morpholinoethane-sulfonic acid (MES), pH 6.6, 40 mM EDTA, 0.02% Tween 20, 7% dimethyl sulfoxide, 20 μg of herring sperm DNA (Promega), 100 μg of bovine serum albumin (BSA), the recommended amount of eukaryotic hybridization control (Affymetrix), and control oligonucleotide B2 for the alignment signal (Affymetrix). The array was automatically washed and stained with streptavidin-phycoerythrin (Invitrogen), using a GeneChip Fluidics Station 450XP (Affymetrix). The probe array was scanned with a GeneChip Scanner 3000 (Affymetrix).

(ii) *Genome-wide expression analysis of the time course experiment.* The expression intensity of each gene at each time point for the three wild-type strains was summarized and scaled by the MAS 5.0 algorithm, using the GeneChip Operating Software (Affymetrix). The intensity datasets for each time point were normalized by the following three steps, using the GeneSpring GX 7.3.1 program (Agilent Technologies, Santa Clara, CA, USA): data transformation (shifting of low signals < 0.01 to 0.01), global scaling (normalization to the median of each array), and normalization to the median of each gene. The microarray data used in this analysis have been deposited in the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/), and the data from the TT broth and synthetic medium analyses are accessible through the GEO Series Accession Nos. GSE10368, and GSE11671, respectively.

(iii) *Differential expression analysis of the Δ*sdrP *strain relative to the wild type.* The image data for the four wild-type and four Δ*sdrP* strains were processed by following the same

procedure as that described above, with the exception that the normalization of each gene was performed using the wild-type data as the control sample. The false discovery rate (*q*-value) (Storey, 2002) of the observed differences in the normalized intensities between the wild-type and Δ*sdrP* strains was calculated using the R program (http://www.r-project.org). The microarray data used in this analysis have been deposited in GEO, and are accessible through the GEO Series Accession No. GSE10369.

(iv) *Differential expression analysis of various stress responses.* For the analysis, 306 datasets from 117 experimental conditions were used (Table 2). The raw intensity data were summarized and scaled by basically the same procedure as that described in section 1-3 (ii). The datasets were normalized through the following normalization steps, using the Subio Platform (Subio, Tokyo, Japan); i.e., shifting of low signals < 1.0 to 1.0, log-based transformation, and global scaling [normalized to the 75th percentile (third quartile) of each array]. The data for physically or chemically treated cells were normalized using the data for the non-treated cells as a control. The *q*-value was calculated using the R program. The microarray data used in this analysis have been summarized in the dataset of the expression pattern analysis (see below).

(v) *Expression pattern analysis.* The details of the procedure are available in the GEO, and the Accession Nos. are provided in Table 2. The raw intensity data were summarized and scaled by essentially the same procedure as that described in section 1-3 (ii). The datasets were

normalized by basically the same procedure as that for the stress response experiment, with the exception that the normalization to the mean value for each gene was performed after the global scaling. The Spearman's correlation coefficients between the *sdrP* gene and each gene were calculated using the Subio Platform (Subio). The microarray data used in this analysis have been summarized and deposited in the GEO, and are accessible through the GEO series Accession No. GSE21875.

Table 2. GEO Accession No. and experimental conditions for the DNA microarray dataset used for expression pattern analysis

GEO Accession No.[¶]	Strain[*]	Culture medium[†]	Experimental type[‡]	Experimental conditions	Number of experimental conditions	Number of samples	References
GSE21290	WT	TT	time course	cultured for 180, 240, 300, 360, 420, 480, 540, 600, 680, and 760 min	10	30	(Agari *et al.*, 2008a)
GSE21473	WT	TT	time course[a]	cultured for 180, 240, 300, 360, 420, 480, 540, 600, 680, and 760 min	10	16	(Shinkai *et al.*, 2007a)
GSE19839	WT	TT	time course	cultured for 180, 360, 540, 680, 880, 1000, 1120, and 1240 min	8	20	
(GSE19747)	WT	SM	time course	cultured for 360, 480, 600, 720, 840, 960, 1080, 1200, 1320, and 1440 min	10	19	
GSE21183	WT	SM	time course	cultured for 600 and 1200 min	2	6	(Agari *et al.*, 2008a)
GSE19723	WT	TT	low-temperature treatment	cooling on ice or with cold ethanol when the cells were harvested	2	6	
GSE19508	WT	SM	$CuSO_4$ treatment	with 1.25 mM $CuSO_4$ for 0 and 30 min	2	6	(Sakamoto *et al.*, 2010)
GSE20900	WT	SM	$ZnSO_4$ treatment	with 1.25 mM $ZnSO_4$ for 0 and 30 min	2	6	(Sakamoto *et al.*, 2010)
GSE21199	WT	SM	$FeSO_4$ treatment	with 1 mM $FeSO_4$ for 0 and 30 min	2	6	
(GSE21466)	WT	SM	$FeSO_4$ treatment	with 1 mM $FeSO_4$ for 30 min	1	3	
GSE21430	WT	TT	H_2O_2 treatment	with 10 mM H_2O_2 for 0, 5, 15, and 30 min	4	12	
GSE21432	WT	TT	tetracycline treatment	with 50 mM tetracycline for 0 and 10 min	2	6	
GSE21434	WT	TT	tetracycline treatment	with 10 mM tetracycline for 5 min	1	3	
GSE21289	WT	TT	NaCl treatment	with 0.5 or 1.5% NaCl for 0 and 30 min	3	9	
GSE21433	WT	SM	diamide treatment	with 10 mM diamide for 0 and 10 min	2	6	
GSE21435	WT	TT	ethanol treatment	with 5% ethanol for 30 min	1	3	
GSE21288	WT	TT	heat treatment	at 80 ºC for 0 and 30 min	2	6	
GSE19759	WT	TT	phage infection	infection of *Φ*YS40 for 0, 25, 50, 75, and 100 min	5	15	(Agari *et al.*, 2010a)
GSE21474	Δ*crp*	TT	phage infection	infection of *Φ*YS40 for 0, 75, and 100 min	3	9	(Agari *et al.*, 2010a)
GSE19509	Δ*csoR*	SM	$CuSO_4$ treatment	with 1.25 mM $CuSO_4$ for 0 and 30 min	2	6	(Sakamoto *et al.*, 2010)
GSE21470	Δ*TTHA0995*	TT	$FeSO_4$ treatment	with 1 mM $FeSO_4$ for 0 and 30 min	2	6	
GSE21195	WT, Δ*TTHA0175*	TT	low-temperature treatment	cultured at 45 ºC for 0, 30, and 120 min	6	12	
(GSE21471)	Δ*TTHA0175*	TT	gene disruption	cultured for 300, 480, and 600 min	3	6	
GSE21617	Δ*TTHA0167*	TT	gene disruption	cultured for 480 min	1	3	

Continued on next page.

Table 2. Continued.

GEO Accession No. [¶]	Strain[*]	Culture medium[†]	Experimental type[‡]	Experimental conditions	Number of experimental conditions	Number of samples	References
GSE21196	Δ*TTHA0344*	TT	gene disruption	cultured for 680 min	1	3	
GSE21197	Δ*TTHA0995*	TT	gene disruption	cultured for 680 min	1	3	
(GSE21456)	Δ*TTHA0252*	TT	gene disruption	cultured for 360, 540, and 680 min	3	6	
GSE21186	Δ*TTHA0655*	TT	gene disruption	cultured for 360 and 680 min	2	6	
(GSE21461)	Δ*TTHA0845*	TT	gene disruption[a]	cultured for 420 min	1	3	
GSE21185	Δ*TTHA0973*	TT	gene disruption	cultured for 680 min	1	3	
GSE21453	Δ*TTHA0101*	TT	gene disruption	cultured for 680 min	1	3	
(GSE21462)	Δ*TTHA1431*	TT	gene disruption[a]	cultured for 360 min	1	3	
(GSE21198)	Δ*TTHA1438*	TT	gene disruption	cultured for 240, 360, 480, 600, and 760 min	5	9	
(GSE21460)	Δ*TTHA1634*	TT	gene disruption	cultured for 420, and 600 min	2	4	
GSE21184	Δ*TTHB099*	TT	gene disruption	cultured for 360, and 420 min	2	8	
(GSE21457)	Δ*TTHB186*	TT	gene disruption[a]	cultured for 360, and 480 min	2	6	
GSE21464	Δ*TTHA1939*	SM	gene disruption	cultured for 800 min	1	4	(Ooga *et al.*, 2009)
GSE21436	Δ*sigE*	TT	gene disruption[a]	cultured for 480 min	1	3	(Shinkai *et al.*, 2007b)
GSE19758	Δ*asiE*	TT	gene disruption	cultured for 420, and 480 min	2	6	(Sakamoto *et al.*, 2008)
GSE21459	Δ*crp*	TT	gene disruption[a]	cultured for 360, and 480 min	2	6	(Shinkai *et al.*, 2007a)
GSE21472	Δ*sdrP*	TT	gene disruption	cultured for 680, 1120, and 1240 min	3	10	(Agari *et al.*, 2008a)
Total					117	306	

[¶]Datasets denoted in bold letters are from GEO; datasets denoted in parentheses are from unpublished data obtained by Aiko Kashihara, Emi Ishido-Nakai, and Miwa Ohmori of RIKEN SPring-8 Center. I reanalyzed these datasets for this study.

[*]Gene-disruptant strains were constructed basically as described previously (Hashimoto *et al.*, 2001). WT, wild type. The references for Δ*crp*, Δ*csoR*, Δ*sigE*, Δ*asiE*, and Δ*sdrP* are (Shinkai *et al.*, 2007a), (Sakamoto *et al.*, 2010), (Shinkai *et al.*, 2007b), (Sakamoto *et al.*, 2008), and (Agari *et al.*, 2008a).

[†]TT, rich medium; SM, synthetic medium. Their compositions are given in the main text.

[‡a], The 3′-terminals of the fragmented cDNAs were labeled using reagent from ENZO Life Sciences.

***In vitro* transcription assays**

Preparation of templates. The upstream regions of the *TTHA0337*, *TTHA1028*, *TTHA0654*, *TTHA0425*, *TTHA0634*, *TTHA0570*, *TTHA0029*, *TTHA0557*, *TTHA1128*, *TTHA1215*, *TTHA1625*, *TTHA1635*, *TTHA1892*, *TTHB132*, *and TTHA0987* genes were amplified by genomic PCR using the primers listed in Table 1. The amplified fragments were digested with BamHI and EcoRI and cloned into pUC19 (Merck). For the construction of plasmids containing the upstream regions of the *TTHA0986* and *TTHA0770* genes, oligonucleotides were annealed, and the partially duplex oligonucleotides were extended by incubation with *E. coli* DNA polymerase I Klenow fragment (New England BioLabs, Ipswich, MA, USA) using oligonucleotides P15 and P16, and P17 and P18, respectively (Table 1), as described previously (Shinkai *et al.*, 2001). Using each plasmid as the template, PCR was performed with primers P19 and P20 (Table 1) to prepare template DNA for the transcription assay (Fig. 7).

Various lengths of template DNA containing the upstream regions of the *TTHA0337* and *TTHA0634* genes were prepared by PCR using the primers listed in Table 1 and plasmids carrying the upstream regions of these genes as templates. The amplified fragments were excised from a 0.8% agarose gel, extracted with phenol and ether, and precipitated with ethanol. The DNA fragments were used for the following assay.

Run-off transcription. Assays were performed in 15 μl of reaction mixtures in the absence or presence of 2 μM *T. thermophilus* SdrP by basically following the same process as that

described previously (Shinkai *et al.*, 2007a). The template DNA was pre-incubated with or without SdrP at 55°C for 5 min. *T. thermophilus* RNAP was added, and the mixture was further incubated for 5 min. Transcription was initiated by the addition of 1.5 μCi [α-^{32}P]-CTP and unlabeled ribonucleotide triphosphates. After further incubation for 10 min, the reaction was stopped, and the sample was analyzed on a 10% polyacrylamide gel containing 8 M urea followed by autoradiography.

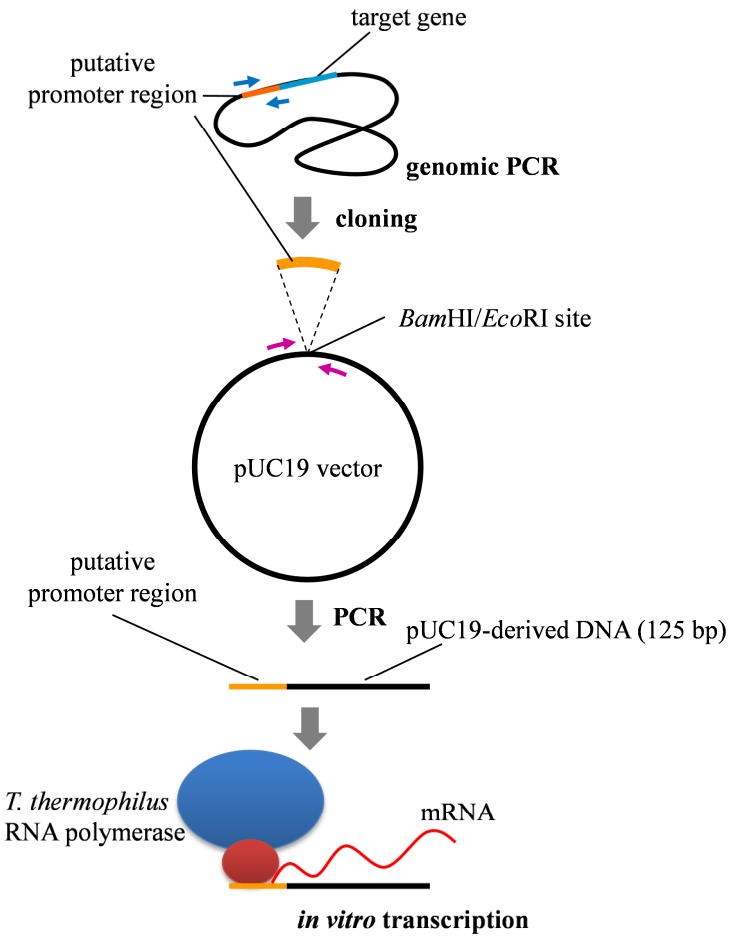


Fig. 7. Schematic diagram of the *in vitro* run-off transcription assay.

Identification of the transcriptional start site

Primer extension analysis with RNA transcribed *in vitro* was performed by basically following the same method as that described previously (Shinkai *et al.*, 2007a) with the exception that 2 μM SdrP was added to the reaction mixture instead of 1 μM CRP and 2 μM cAMP. The nucleotide sequence of the template DNA was determined by the dideoxy-mediated chain termination method (Sanger *et al.*, 1992). Samples were analyzed on an 8% polyacrylamide gel containing 8 M urea followed by autoradiography.

RT-PCR analysis

The *T. thermophilus* wild-type and *csoR* gene-deficient (Δ*csoR*) strains (Sakamoto *et al.*, 2010) were cultured at 70°C in a rich or synthetic medium (Table 2). The details of the culture conditions are given in the GEO, the Accession Nos. being GSE21433 (for diamide treatment), GSE21430 (for H_2O_2 treatment), GSE20900 (for $ZnSO_4$ treatment), GSE21432 (for tetracycline treatment), GSE21289 (for NaCl treatment), GSE21435 (for ethanol treatment), GSE19508 (for $CuSO_4$ treatment of the wild-type strain), and GSE19509 (for $CuSO_4$ treatment of the Δ*csoR* strain). Total RNA was isolated from each strain according to the procedure described in the "DNA microarray analysis" subsection. Using the RNA (1 μg) as a template, RT-PCR was performed in 20 μl reaction mixtures with a PrimeScript RT-PCR kit (Takara Bio, Shiga, Japan) according to the manufacturer's instructions. The reverse transcription reaction was performed

at 42°C for 20 min. Using 1 μl of the reaction mixture as a template, PCR was performed in the presence of 0.2 μM each of the forward and reverse primes in a 25 μl reaction mixture. After the reaction, samples were analyzed on a 2% agarose gel, followed by staining with ethidium bromide and photography. The primers used are listed in Table 1.

Other methods

N-terminal sequence analysis of proteins was performed with a protein sequencer (Procise HT; Applied Biosystems, Carlsbad, CA, USA). Dynamic light scattering photometry was performed with a DynaPro-801 detector (Wyatt Technology, Santa Barbara, CA, USA). BLAST and CDD searches were performed on the http://blast.ncbi.nlm.nih.gov/Blast.cgi, and http:// www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi websites, respectively.

1-4. Results

Amino acid sequence of the *T. thermophilus* SdrP

The *sdrP* ORF (*TTHA1359*) encodes 202 amino acid residues (NCBI Accession No. YP_144625) with a predicted molecular mass of 22,320 Da. Based on the results of a conserved domain database (CDD) (Marchler-Bauer *et al.*, 2002) search, the protein has two conserved domains. One is a cyclic nucleotide monophosphate-binding domain (cd00038.3) comprising residues E5–R79 with an *e*-value of 3e-08 for the consensus sequence, and the other is a helix-turn-helix DNA-binding domain (smart 00419.11) comprising residues V142–D186 with an *e*-value of 2e-05 for the consensus sequence (Fig. 8). A BLAST search revealed that the SdrP protein was identical to the TTC0994 protein from *T. thermophilus* HB27 (*e* = 4e-109). Homologous proteins that were most closely related to this protein were present in *Deinococcus geothermalis* DSM 11300 (9e-45), *Deinococcus radiodurans* R1 (2e-37), *T. thermophilus* HB27 (2e-21), *T. thermophilus* HB8 (2e-21) and marine actinobacterium PHSC20C1 (2e-16). These homologues belong to the CRP/FNR family (Fig. 8). Among the proteins with known functions, *Bacillus licheniformis* ArcR (Maghnouj *et al.*, 2000; Wöhlkonig *et al.*, 2004), *Synechocystis* sp. PCC 6803 NtcA (Alfonso *et al.*, 2001; Harano *et al.*, 1997), *Synechococcus elongates* PCC7942 NtcA (Vázquez-Bermúdez *et al.*, 2002), *Streptomyces coelicolor* CRP (Derouaux *et al.*, 2004), *T. thermophilus* CRP (Shinkai *et al.*, 2007a) and *E. coli* CRP (Kolb *et al.*, 1993) exhibited *e*-values of 1e-14, 1e-12, 3e-12, 1e-11, 1e-10, and 1e-9, respectively.

Cysteine residues that are found in several CRP/FNR family proteins that sense oxygen or redox variations (Green *et al.*, 2001; Kiley & Beinert, 2003) were not present in the SdrP (Fig. 8).

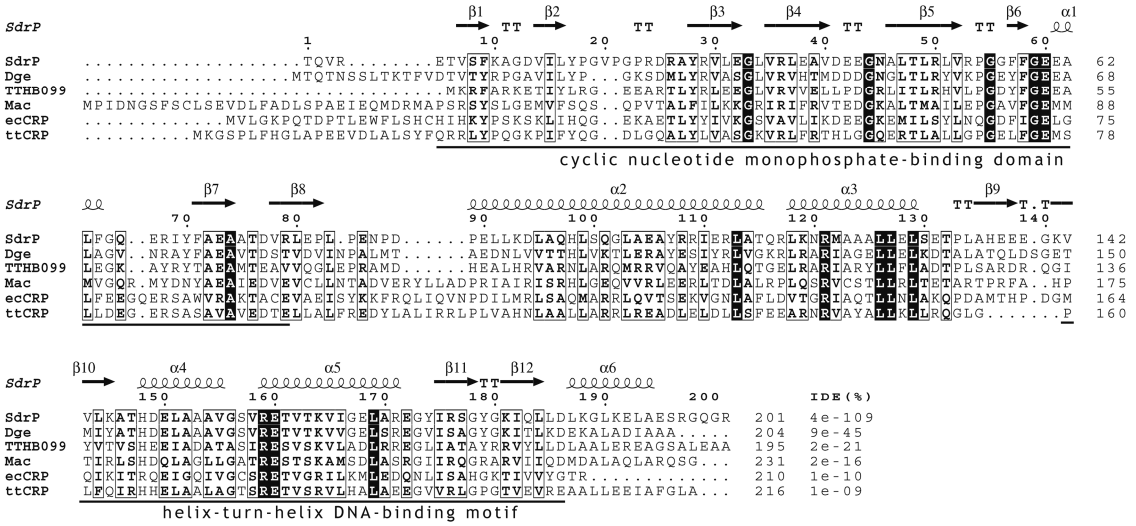


Fig. 8. Sequence alignment of *T. thermophilus* SdrP with representative homologous proteins. Strictly conserved residues are represented by white letters on a black background, and similar residues are depicted by boxed bold letters. Dge, *D. geothermalis* DSM11300 Dgeo_1015 (YP_604484); TTHB099, *T. thermophilus* HB8 TTHB099 (YP_145338); Mac, marine actinobacterium PHSC20C1 A20C1_09144 (ZP_01129038); ecCRP, *E. coli* CRP (Kolb *et al.*, 1993); and ttCRP, *T. thermophilus* HB8 CRP (Shinkai *et al.*, 2007a). The sequences were aligned using ClustalwW2 (Larkin *et al.*, 2007). The secondary structure was calculated from crystal structure of SdrP using DSSP (Kabsch & Sander, 1983), and the figure was generated with ESPript 2.2 (Gouet *et al.*, 1999). The percentage identities (IDE) to SdrP is indicated on the right

Initial characterization of recombinant *T. thermophilus* SdrP

The *sdrP* gene was overexpressed in *E. coli*, and the recombinant protein was purified from the cell lysate. The lysate, which was resistant to treatment at 70°C for 10 min, was fractionated on a hydrophobic column. This was followed by ammonium sulfate precipitation and further fractionation by anion-exchange, hydroxyapatite and gel-filtration column chromatographies. Protein with purity > 95% on SDS-polyacrylamide gel electrophoresis (PAGE) was obtained (data not shown). The N-terminal amino acid sequence of the purified

protein was TQVRE, indicating that the N-terminal methionine had been deleted (data not shown). The molecular mass of SdrP as estimated by light scattering photometry was 48.9 kDa (data not shown), suggesting that *T. thermophilus* SdrP exists as a homodimer in solution.

Crystal structure of *T. thermophilus* SdrP

The SdrP crystals grew within 3 days to maximum dimensions of 0.2 × 0.2 × 0.2 mm (Fig. 9). The crystal structure of Se-SdrP was determined by the multiple wavelength anomalous dispersion method and refined to 1.5 Å resolution. The data collection, model and refinement statistics are summarized in Table 3. The overall structure of *T. thermophilus* SdrP is shown in Fig. 10A. The final model comprises residues 5–198 of 201 together with 141 water molecules. Residues 1–4 and 199–201 are not included in the model due to their poor electron density. The N-terminal domain of *T. thermophilus* SdrP consists of one α-helix (α1, residues 61–64) and eight β-strands (β1, residues 7–9; β2, residues 14–16; β3, residues 28–32; β4, residues 35–40; β5, residues 46–52; β6, residues 57 and 58; β7, residues 71–74; and β8, residues 78–82) that adopt a double-stranded β-helix fold with a jelly roll topology. The C-terminal domain of *T. thermophilus* SdrP consists of four α-helices (α3, residues 118–129; α4, residues 148–155; α5, residues 159–171; and α6,

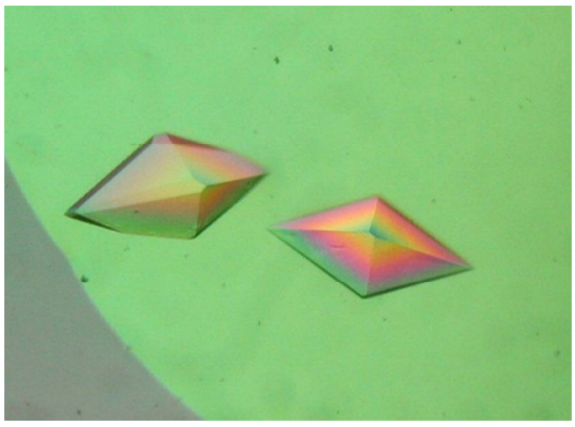


Fig. 9. Crystals of *T. thermophilus* SdrP.

residues 187–194) and four β-strands (β9, residues 135–138; β10, residues 141–145; β11, residues 175–178; and β12, residues 181–184) that adopt a winged helix-turn-helix fold. These two domains are connected by a large α-helix (α2, residues 89–115) as the linker.

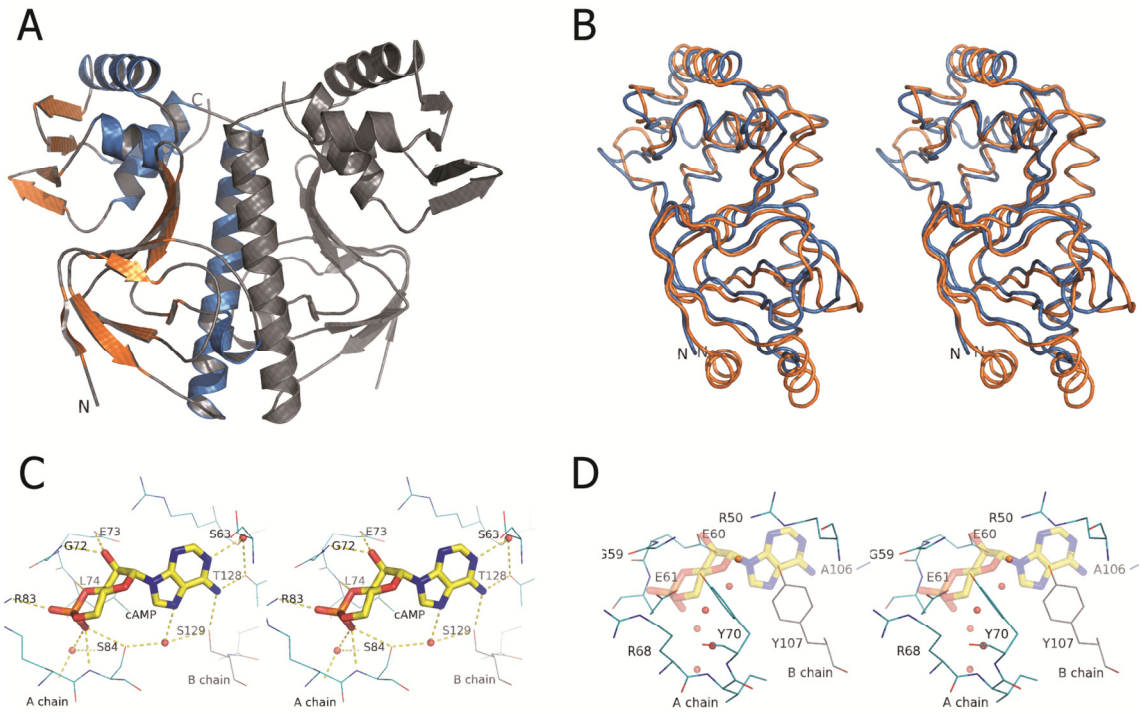


Fig. 10. (A) Ribbon diagram of the *T. thermophilus* SdrP dimer. The dimer molecule was generated around a crystallographic twofold axis. The α-helices and β-strands in one chain are coloured blue and orange, respectively, and the other chain is coloured dark grey. **(B)** Stereoview of superpositioning of the main chain structure of *T. thermophilus* SdrP (blue) and the DNA-binding form of *E. coli* CRP (orange) (Passner and Steitz, 1997). **(C)** Stereoview of the primary cAMP binding site of *E. coli* CRP. Residues involved in cAMP binding are labeled. **(D)** Stereoview of the site in *T. thermophilus* SdrP corresponding to the primary cAMP binding site in *E. coli* CRP shown in (C). A cAMP molecule of the *E. coli* CRP-cAMP-DNA complex is superimposed on SdrP as a transparent stick model. These figures were drawn using the PyMol program (http://pymol.sourceforge.net).

Table 3. X-ray data collection and refinement statistics

	Remote	Peak	Edge
Data collection			
Wavelength (Å)	0.9000	0.9787	0.9794
Resolution (Å)	50-1.45 (1.50-1.45)	50-1.50 (1.55-1.50)	50-1.50 (1.55-1.50)
Space group		$P4_12_12$	
No. of molecules in an asymmetric unit		1	
Unit cell parameters (Å, °)		a = 54.05, b = 54.05, c = 147.718	
		α = β = γ = 90	
No. of measured reflections	521,589	432,807	433,150
No. of unique reflections	39,799	35,176	35,240
Completeness (%)	99.6 (96.1)	97.4 (81.4)	97.4 (81.1)
Redundancy	13.1 (8.4)	12.3 (6.6)	12.3 (6.6)
I/σ(I)	50.7 (7.7)	44.9 (9.7)	59.8 (7.9)
R_{merge}[a](%)	4.6 (26.6)	4.8 (18.9)	4.2 (22.9)
Refinement			
Resolution (Å)	36.4-1.5		
R_{work}[b](%)/R_{free}[c](%)	22.1/23.2		
No. of protein atoms/water atoms	1,507/141		
RMSD bond lengths (Å)	0.005		
RMSD bond angles (°)	1.3		
Ramachandran plot (%)			
Most favored	93.4		
Allowed	6.6		
Disallowed	0.0		

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 i[th] 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.

The overall structure was compared with previously determined structures in the PDB database using the DALI server (Holm & Sander, 1998). The closest structure was that of *E. coli* CRP complexed with cAMP and DNA (PDB code: 2CGP chain A) (Passner & Steitz, 1997), and the *Z*-score and root mean square deviation (r.m.s.d.) were 20.1 and 2.3 Å, respectively (Fig. 10B). SdrP does not have an N-terminal α-helix corresponding to that found in *E. coli* CRP. Instead, it has a C-terminal α6 helix that is not found in *E. coli* CRP (Fig. 10B). Superpositioning of the structure of SdrP with that of *E. coli* CRP revealed that residues G72, E73, L74, R83, S84, T128 and S129 of *E. coli* CRP (Fig. 8), which are primary cAMP binding sites (Passner & Steitz, 1997), correspond to G59, E60, E61, R68, Y70, A106 and Y107 in SdrP (Fig. 10C and D). It should be noted that the side-chains of E60, E61, Y70 and Y107 of SdrP penetrate into a space corresponding to the cAMP-binding pocket of *E. coli* CRP (Fig. 10D). The conformation of residues A46, G156 and R159 of SdrP is similar to that of the corresponding residues of *E. coli* CRP (E59, G178 and R181) involved in secondary cAMP binding (Passner & Steitz, 1997). No small molecules except MPD, which was present in the solution used for crystallization, were found in the crystal structure of SdrP.

The other structural homologues of SdrP that are closely related include the transcriptional regulator PrfA from *Listeria monocytogenes* [PDB code: 1OMI chain A, $Z = 14.8$, r.m.s.d. = 3.0 Å, sequence identity (IDE) = 15%], the regulatory subunit of cAMP-dependent protein kinase from *Bos taurus* (bovine) (PDB code: 1RGS, $Z = 11.4$, r.m.s.d. = 3.5 Å, IDE = 27%), the cyclic

nucleotide-binding domain of an ion channel from *Rhizobium loti* (PDB code: 1PF0, Z = 11.0, r.m.s.d. = 2.5 Å, IDE = 33%), the regulatory domain of guanine nucleotide exchange factor Epac2 from *Mus musculus* (PDB code: 1O7F, Z =10.4, r.m.s.d. = 5.3 Å, IDE = 18%), transcription factor CooA from *Rhodospirillum rubrum* (PDB code: 1FT9, Z = 10.3, r.m.s.d. = 5.1 Å, IDE = 18%), and the oxidized form of transcription factor CprK from *Desulfitobacterium hafniense* complexed with *o*-chlorophenolacetic acid (PDB code: 2H6B, Z = 10.2, r.m.s.d. = 9.8 Å, IDE = 13%). The atomic coordinates and structure factors have been deposited in the Protein Data Bank under PDB ID code 2ZCW.

Effects of disruption of the *T. thermophilus sdrP* gene

Using Affymetrix GeneChip microarrays, I examined the expression profile of *T. thermophilus sdrP* mRNA *in vivo* during cultivation at 70°C. In a rich medium, the normalized intensity of *sdrP* mRNA in the logarithmic phase (5 h cultivation, A_{600} = ~0.7) was 0.53 ± 0.05 while that in the stationary phase (11.3 h cultivation, A_{600} = ~4.9) was 11.6 ± 3.03, indicating that the expression level increased by more than 20-fold (Fig. 11A). Increased expression of the *sdrP* mRNA in the stationary phase was also observed in a synthetic medium, i.e. the normalized intensities in the logarithmic (10 h cultivation, A_{600} = ~0.5) and stationary (20 h cultivation, A_{600} = ~4.6) phases were 0.42 ± 0.06 and 3.63 ± 2.59 respectively.

To determine the effects of *T. thermophilus* SdrP *in vivo*, I disrupted the *sdrP* gene of the *T.*

thermophilus HB8 strain and compared its growth with that of the wild type. The *sdrP*-disruptant (Δ*sdrP*) strain was viable, indicating that this gene is not essential for this strain (Fig. 11A and B). In comparison with the wild type, the Δ*sdrP* strain exhibited a slight growth defect and reached the stationary phase at a lower cell density during cultivation in a rich medium (Fig. 11A). When the strains were cultivated in a synthetic medium, the growth defect of the Δ*sdrP* strain was more significant even in the logarithmic phase (Fig. 11B). Furthermore, I found that in comparison with the wild type, the growth of the Δ*sdrP* strain was more significantly affected by diamide treatment (Fig. 11B).

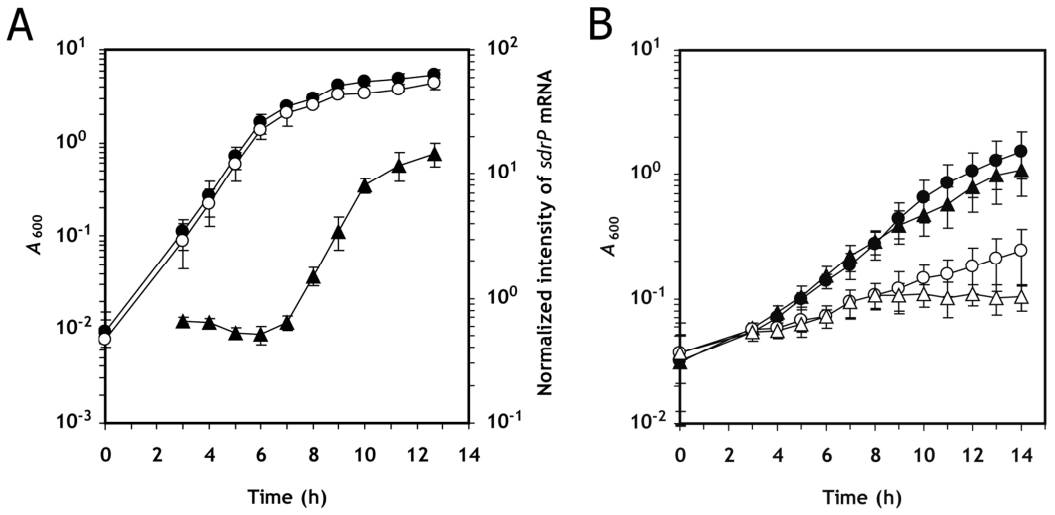


Fig. 11. (A) Three clones of the wild type (●) and Δ*sdrP* (○) strains of *T. thermophilus* HB8 were individually grown in a rich medium, and the A_{600} values at the indicated times are expressed as means ± SD respectively. The expression of *sdrP* mRNA (▲) in the wild-type strain was investigated at the indicated times using GeneChip technology and expressed as normalized intensity ± SD. **(B)** Effects of diamide on the growth of the wild-type (● and ▲) and Δ*sdrP* (○ and △) strains of *T. thermophilus* HB8 in a synthetic medium. Cells were grown at 70°C. After 8 h, diamide was either added (▲ and △) or not added (● and ○) to a final concentration of 2 mM. The three clones were individually grown in a synthetic medium, and the A_{600} values at the indicated times are expressed as means ± SD respectively.

Screening of SdrP-regulated genes by means of differential gene expression analysis

Next, to find genes that are regulated by SdrP, four wild-type and four Δ*sdrP* strains were cultured for 680 min (A_{600} = 4–5 for wild type and 2.5–4 for Δ*sdrP*) in rich medium, and a genome-wide gene expression analysis was performed using GeneChip. The expression level of the *sdrP* mRNA in the Δ*sdrP* strain relative to that in the wild type was 0.016 (*q*-value = 0.027), indicating that the *sdrP* gene was disrupted. From the 2,205 genes analyzed, I selected those that showed altered expression in the Δ*sdrP* strain with *q*-values of less than 0.06. In total, 25 genes on the chromosomal DNA (designated as TTHA) and one gene on megaplasmid pTT27 (designated as TTHB) were selected (Table 4). Of the total 26 genes, 16 showed lower levels in the Δ*sdrP* strain relative to those in the wild type (Table 4).

Table 4. Genes exhibiting altered expression in the Δ*sdrP* strain in comparison with the wild type. The expression levels in the Δ*sdrP* strain relative to that in the wild type and the *q*-values of the observed differences between the wild-type and Δ*sdrP* strains are shown. Only genes for which the *q*-values are < 0.06 are shown

Gene Name	Annotation for product	Expression[a] (*q*)
TTHA0425	NAD(P)H oxidase	0.303 (0.009)
TTHA1369	phospholipase domain protein	2.275 (0.014)
TTHA0030	hypothetical protein	0.292 (0.018)
TTHA0638	hypothetical protein	0.409 (0.025)
TTHA0570	Glucose/sorbosone dehydrogenase	0.178 (0.025)
TTHA0637	Uncharacterized protein with a von Willebrand factor type A domain	0.307 (0.027)
TTHA0769	DegQ, Trypsin-like serine protease	0.453 (0.027)
TTHA1102	hypothetical protein	1.840 (0.027)
TTHA0655	Predicted transcriptionalregulator	0.154 (0.027)
TTHA0340	hypothetical protein	1.757 (0.027)
TTHA1810	hypothetical protein	5.378 (0.027)
TTHA0460	MutT/nudix family protein	0.300 (0.027)
TTHA1570	deoxyhypusine synthase	2.177 (0.027)
TTHB243	hypothetical protein	0.398 (0.028)
TTHA0986	Highly conserved protein containing thioredoxin domain	0.074 (0.033)
TTHA0337	hypothetical protein	0.057 (0.043)
TTHA1771	pyrimidine-nucleoside (thymidine) phosphorylase	1.753 (0.043)
TTHA1128	probable peptidase	0.147 (0.054)
TTHA1243	septum site-determining protein MinD	1.603 (0.054)
TTHA1028	SseA, Rhodanese-related sulfurtransferase	0.063 (0.054)
TTHA1176	hypothetical protein	1.893 (0.054)
TTHA0035	hypothetical membrane protein	0.586 (0.054)
TTHA1185	GTP-binding protein	3.040 (0.054)
TTHA0520	NAD-dependent malic enzyme (malate dehydrogenase)	0.462 (0.054)
TTHA1803	pterin-4-alpha-carbinolamine dehydratase	0.216 (0.059)
TTHA1423	cytochrome c-552 precursor	2.220 (0.059)

a (column 3). Normalized intensity of the Δ*sdrP* strain relative to that of the wild type.

Effects of *T. thermophilus* SdrP on transcription

DNA fragments upstream of several genes and gene clusters that showed altered expression in the Δ*sdrP* strain (*q*-value < 0.06) (Tables 4 and 5) were cloned and used as templates for *in vitro* run-off transcription assays. I found that the DNA fragments containing the upstream regions of the *TTHA0986*, *TTHA0770*, *TTHA0337*, *TTHA1028*, *TTHA0654*, *TTHA0425*, *TTHA0634*, and *TTHA0570* genes were transcribed by *T. thermophilus* RNAP holoenzyme containing a housekeeping σ, i.e. σ^A (Vassylyeva *et al.*, 2002; Wnendt *et al.*, 1990), in an SdrP-dependent manner (Fig. 12A). Using GeneChip technology, I investigated the expression profiles of these eight genes during cultivation in a rich medium at 70°C. I found that the expression levels of these genes tended to increase in the stationary phase in comparison with the late-logarithmic phase (8 h cultivation) (Fig. 13A and B). However, their mRNA expression profiles did not necessarily parallel that of *sdrP* mRNA, especially in the case of the *TTHA0425* gene, the highest expression level of which was observed in the early logarithmic phase (~3 h cultivation) (Fig. 13A and B).

Table 5. Genes under the control of *T. thermophilus* SdrP-dependent promoters and their expression in the Δ*sdrP* strain relative to that in the wild type

Gene name	Expression (*q*-value)[a]	Conserved domain	*e*-value	Annotation for product[b]	Possible cellular role	Reference
TTHA0337	0.057 (0.043)			hypothetical protein	unknown	
TTHA0425	0.303 (0.009)	cd03370	1e-29	NAD(P)H oxidase	supply energy; redox control	(Park *et al.*, 1992)
TTHA0570	0.178 (0.025)	COG2133	8e-51	glucose/sorbosone dehydrogenase	supply energy	
TTHA0634	0.566 (0.117)	COG1239	3e-56	magnesium chelatase subunit Chl I	chelation of metal ion to the predicted nucleotidyltransferase, TTHA0635/0636	(Jasiecki & Wegrzyn, 2003; Lehmann *et al.*, 2003; Lehmann *et al.*, 2005; Martin & Keller, 2007; Santos *et al.*, 2006); PDB code: 1WWP
TTHA0635	0.654 (0.157)	COG1669	3e-05	predicted nucleotidyltransferase	polyadenylation of mRNA	
TTHA0636	0.479 (0.099)	pfam08780	3e-25	nucleotidyltransferase substrate-binding protein-like protein	polyadenylation of mRNA	
TTHA0637	0.307 (0.027)	COG4867	3e-78	uncharacterized protein with a von Willebrand factor type A domain	unknown	
TTHA0638	0.409 (0.025)			hypothetical protein	unknown	
TTHA0654	0.109 (0.074)	cd02037	6e-52	MRP (multiple resistance and pH adaptation)-like protein	supply energy; redox control	(Almeida *et al.*, 2005)
TTHA0655	0.154 (0.027)	COG2345	4e-16	predicted transcriptional regulator	unknown	
TTHA0769	0.453 (0.027)	COG0265	4e-20	DegQ, trypsin-like serine protease	protein quality control; supply nutrient	(Gottesman, 2003; Kim & Kim, 2005; Reeve *et al.*, 1984; Tsilibaris *et al.*, 2006; Watanabe *et al.*, 1999)
TTHA0770	0.567 (0.196)	COG0466	1e-180	ATP-dependent Lon protease	protein quality control; supply nutrient	
TTHA0986	0.074 (0.033)	COG1331	3e-154	highly conserved protein containing a thioredoxin domain	redox control	
TTHA1028	0.063 (0.054)	COG2897	4e-83	SseA, rhodanese-related sulfurtransferase	redox control	(Nandi *et al.*, 2000); PDB code: 1UAR

a (column 2). Normalized intensity of the Δ*sdrP* strain relative to that of the wild type.
b (column 5). A product that was not found in the CDD search was annotated as a hypothetical protein.
The possible cellular role of the gene product is shown together with the name of the most closely related domain, *e*-value, and annotation of the gene product, which were obtained from the CDD search.

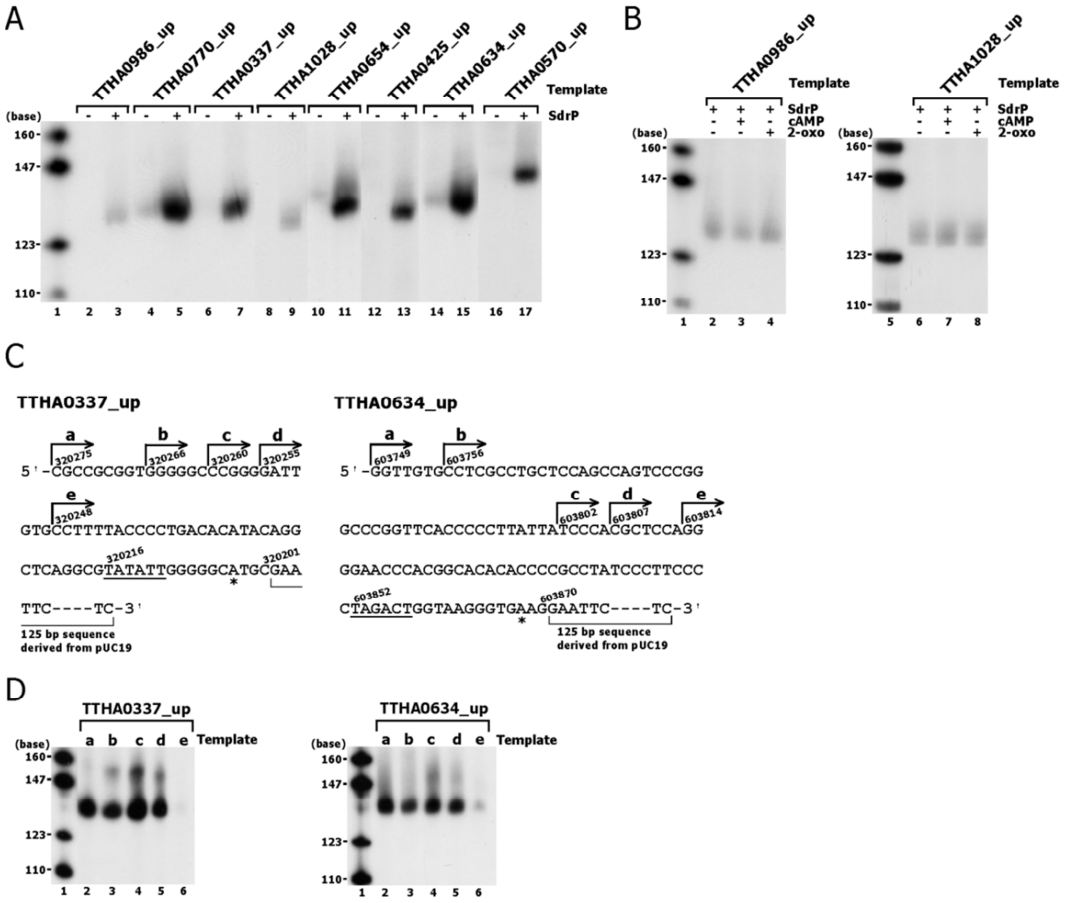


Fig. 12. (A) Run-off transcription assay performed with a template containing the upstream sequence of the *TTHA0986* (TTHA0986_up), *TTHA0770* (TTHA0770_up), *TTHA0337* (TTHA0337_up), *TTHA1028* (TTHA1028_up), *TTHA0654* (TTHA0654_up), *TTHA0425* (TTHA0425_up), *TTHA0634* (TTHA0634_up), or *TTHA0570* (TTHA0570_up) gene in the absence (-) or presence (+) of *T. thermophilus* SdrP. After the reaction, equivalent volumes of samples were analyzed by PAGE followed by autoradiography. Lane 1, [α-^{32}P]-dCTP-labelled MspI fragments of pBR322. **(B)** Run-off transcription assay, performed with TTHA0986_up and TTHA1028_up as the templates in the presence of *T. thermophilus* SdrP (lanes 2 and 6), both SdrP and 2 mM cAMP (lanes 3 and 7), or both SdrP and 2 mM 2-oxoglutalate (2-oxo) (lanes 4 and 8). After the reaction, equivalent volumes of samples were analyzed by PAGE followed by autoradiography. Lanes 1 and 5, [α-^{32}P]-dCTP-labelled MspI fragments of pBR322. **(C)** Upstream sequences of the *THA0337* and *TTHA0634* genes, which were used as the templates for the run-off transcription assays. The numerals represent the genome positions in chromosomal DNA. The transcriptional start sites are indicated by asterisks. Possible -10 hexamer sequences are underlined. **(D)** Run-off transcription assay performed with templates a, b, c, d, and e of TTHA0337_up and TTHA0634_up shown in (C) in the presence of *T. thermophilus* SdrP. After the reaction, equivalent volumes of samples were analyzed by PAGE followed by autoradiography. Lane 1, [α-^{32}P]-dCTP-labelled MspI fragments of pBR322.

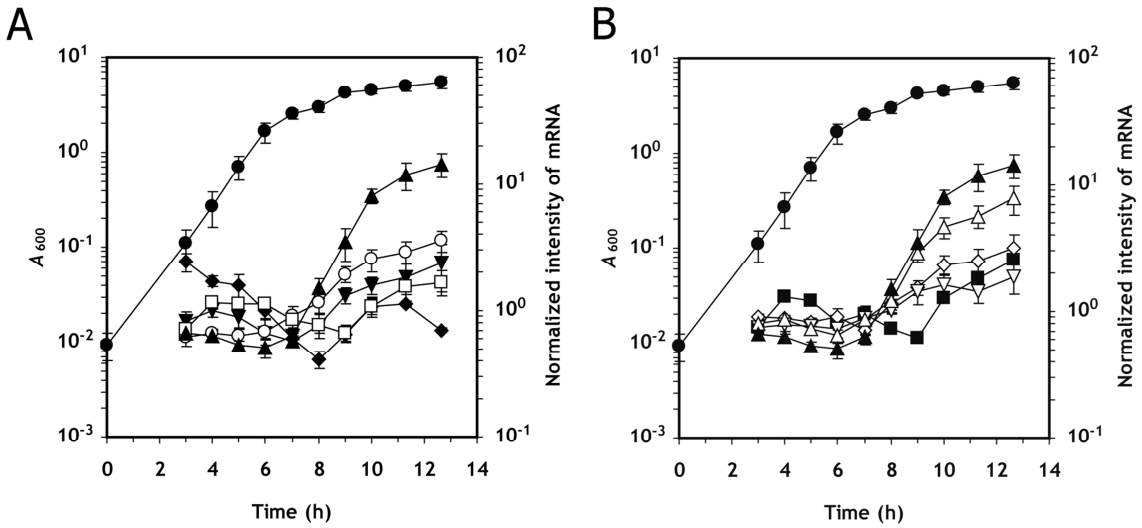


Fig. 13. (A and B) Three clones of the wild-type strain were individually grown in a rich medium, and the A_{600} values at the indicated times are expressed as means ± SD (●). The expression of *TTHA0770* (○), *TTHA0654* (▼), *TTHA1028* (□), *TTHA0425* (◆), *TTHA0986* (△), *TTHA0570* (◇), *TTHA0337* (■), *TTHA0634* (▽) and *sdrP* (▲) mRNAs in the wild-type strain was investigated at the indicated times using GeneChip technology, and expressed as normalized intensity ± SD.

Cyclic AMP and 2-oxoglutarate, which are known effector molecules for CRP (Kolb *et al.*, 1993) and NtcA (Tanigawa *et al.*, 2002; Vázquez-Bermúdez *et al.*, 2002), respectively had no effect on SdrP activity (Fig. 12B). No effect of SdrP was observed with the DNA fragment upstream of the *TTHA1102*, *TTHA1369*, *TTHA1771*, or *TTHA1811* (upstream of *TTHA1810* in the same operon) gene, which showed increased expression in the Δ*sdrP* strain (data not shown). The DNA fragment containing the promoter region of the *sdrP* gene was transcribed with the RNAP-σ^A holoenzyme; however, the transcription was not altered in the presence of SdrP (data not shown). *T. thermophilus* RNAP holoenzyme containing a sole alternative σ, i.e. σ^E (Shinkai *et al.*, 2007b), did not transcribe the above-described genes either in the presence or in the absence of SdrP (data not shown).

To determine the regions necessary for SdrP-dependent transcription, we constructed DNA

templates of various lengths for *in vitro* transcription assays (Fig. 12C). In the case of the *TTHA0337* gene, SdrP-dependent transcription occurred to the same extent even when sequences upstream of position 320,255 were deleted; however, the transcription efficiency was dramatically reduced when sequences upstream of position 320,248 were deleted (Fig. 12D). These results indicate that sequences downstream of position 320,255 are necessary for SdrP-dependent transcription. In the case of the *TTHA0634* gene, the sequences downstream of position 603,807 are necessary for SdrP-dependent transcription (Fig. 12D).

To identify the transcriptional start sites of the SdrP-regulated genes, the transcripts synthesized *in vitro* were reverse transcribed, and their 3′-terminal nucleotides were identified (Fig. 14A). The results indicate that the transcriptional start sites of the genes are 7–11 bp downstream from the predicted -10 hexamers of their promoters (Fig. 15A). Table 5 summarizes the genes that are under the control of the SdrP-dependent promoter found in this analysis and also shows the altered expression levels in the Δ*sdrP* strain relative to those in the wild type.

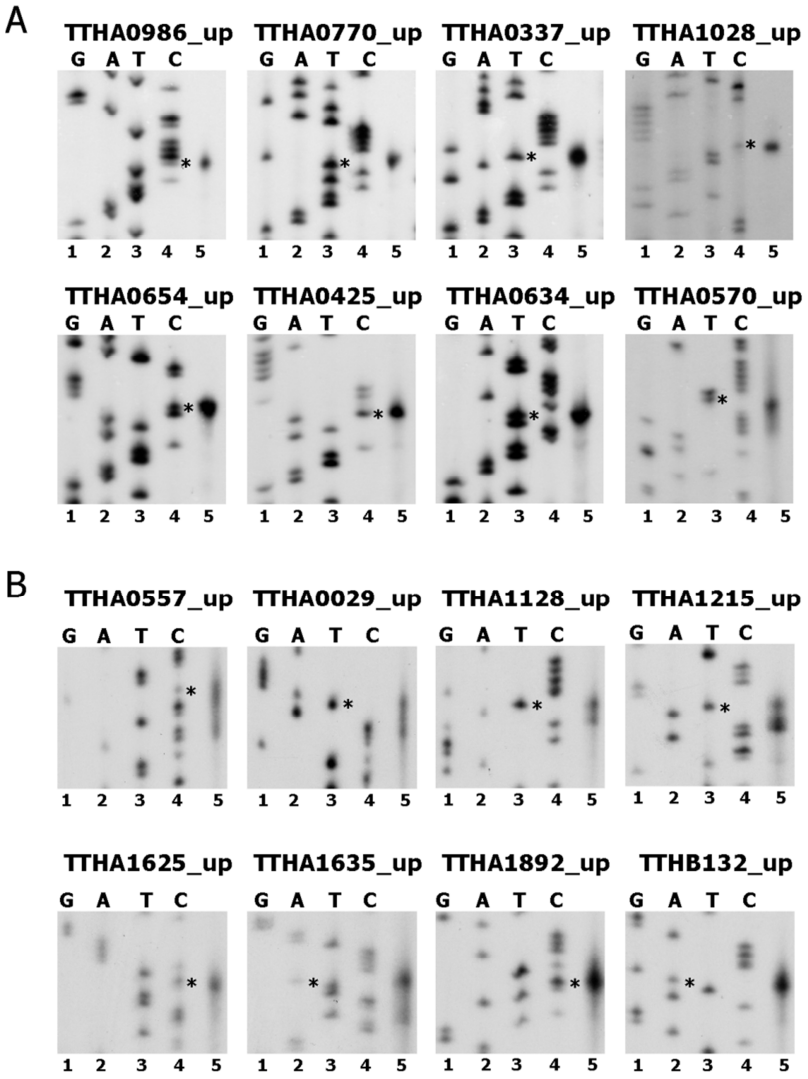


Fig. 14. (A) Identification of the transcriptional start sites of SdrP-regulated genes which were identified by means of differential expression analysis. The RNA transcribed *in vitro* from the gene containing the sequence upstream of *TTHA0986* (TTHA0986_up), *TTHA0770* (TTHA0770_up), *TTHA0337* (TTHA0337_up), *TTHA1028* (TTHA1028_up), *TTHA0654* (TTHA0654_up), *TTHA0425* (TTHA0425_up), *TTHA0634* (TTHA0634_up), or *TTHA0570* (TTHA0570_up) in the presence of *T. thermophilus* SdrP was reverse transcribed (lane 5). The nucleotide sequence of the template DNA was determined by the dideoxy-mediated chain termination method (lanes 1–4). After the reaction, the samples were analyzed by PAGE followed by autoradiography. **(B)** Identification of the transcriptional start sites of SdrP-regulated genes which were identified by means of expression pattern analysis. The RNA transcribed *in vitro* from the gene containing the sequence upstream of *TTHA0557* (TTHA0557_up), *TTHA0029* (TTHA0029_up), *TTHA1128* (TTHA1128_up), *TTHA1215* (TTHA1215_up), *TTHA1625* (TTHA1625_up), *TTHA1635* (TTHA1635_up), *TTHA1892* (TTHA1892_up), or *TTHB132* (TTHB132_up) in the presence of *T. thermophilus* SdrP was reverse transcribed (lane 5). The nucleotide sequence of the template DNA was determined by the same procedure as in (A) (lanes 1–4). The 3′-terminus of the cDNA is indicated by an asterisk.

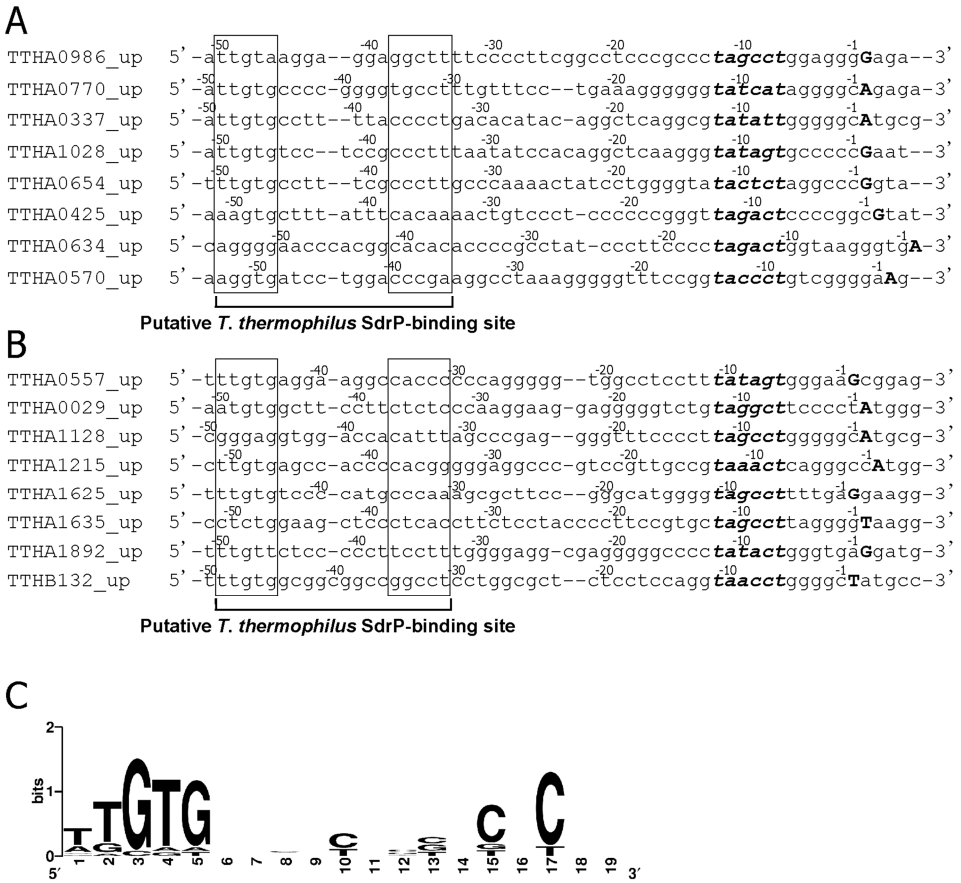


Fig. 15. (A) Nucleotide sequence alignment of predicted SdrP-dependent promoters found in TTHA0986_up, TTHA0770_up, TTHA0337_up, TTHA1028_up, TTHA0654_up, TTHA0425_up, TTHA0634_up, and TTHA0570_up, by means of differential expression analysis of Δ*sdrp* strain. The putative *T. thermophilus* SdrP-binding site are indicated. Possible -10 hexamer sequences of the promoters are indicated by bold italic letters. The *in vitro* transcriptional start sites (see Fig. 14) are indicated by bold capital letters. The numerals represent the positions from the transcriptional start site. **(B)** Nucleotide sequence alignment of predicted SdrP-dependent promoters found in TTHA0557_up, TTHA0029_up, TTHA1128_up, TTHA1215_up, TTHA1625_up, TTHA1635_up, TTHA1892_up, and TTHB132_up, by means of expression pattern analysis are indicated as same as (A). **(C)** Sequence conservation at the putative SdrP-binding site. The sequence logos (Schneider & Stephens, 1990) of the 16 putative SdrP-binding site indicated in (A) and (B) were created by WebLogo (Crooks *et al.*, 2004).

Environmental stresses that induce expression of the *sdrP* gene

The growth of Δ*sdrP* strain was more significantly affected by diamide treatment, in comparison with that of the wild type (Fig. 11B). In order to determine if oxidative stress induces expression of the *sdrP* gene, I treated the wild-type *T. thermophilus* HB8 strain in the

logarithmic phase with diamide or H_2O_2. RT-PCR analysis showed that expression of the *sdrP* gene increased with the addition of a final concentration of 2 mM diamide or 10 mM H_2O_2 (Fig. 16), which was supported by DNA microarray analysis results that showed that expression of the gene increased 27-fold (*q*-value = 0.00) and 10-fold (*q*-value = 0.00) in response to diamide and H_2O_2 treatment, respectively (Table 6). Next, I examined if other environmental or chemical stresses, such as heavy metal ion ($ZnSO_4$ and $CuSO_4$), antibiotic (tetracycline), high salt (NaCl), and organic solvent (ethanol) stresses, induce expression of the *sdrP* gene. RT-PCR (Fig. 16) and DNA microarray (Table 6) analyses indicated that expression of the *sdrP* gene was induced by all of these stresses. In the Δ*csoR* strain, in which excess Cu(I) ions may accumulate due to a significant decrease in the expression of the probable copper efflux P-type ATPase gene *copA* (Sakamoto *et al.*, 2010), the effect of excess $CuSO_4$ on expression of the *sdrP* gene was more significant than that in the wild-type strain (Fig. 16 and Table 6).

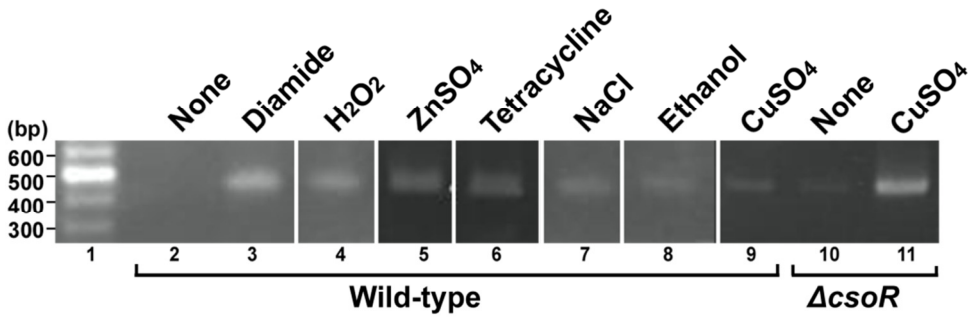


Fig. 16. RT-PCR analysis was performed to detect *sdrP* mRNA for total RNA isolated from the *T. thermophilus* wild-type (lanes 2–9) and Δ*csoR* (lanes 10 and 11) strains cultivated in the absence (lanes 2 and 10) or presence of 2 mM diamide for 30 min (lane 3), 10 mM H_2O_2 for 5 min (lane 4), 1 mM $ZnSO_4$ for 30 min (lane 5), 50 mM tetracycline for 10 min (lane 6), 1.5% NaCl for 30 min (lane 7), 5% ethanol for 30 min (lane 8), or 1.25 mM $CuSO_4$ for 30 min (lanes 9 and 11), respectively, and samples were analyzed on a 2% agarose gel, followed by staining with ethidium bromide and photography. The PCR analysis involved 20 cycles of 98 °C for 1 min, 65 °C for 1 min, and 72 °C for 1 min. Lane 1, 100 bp DNA ladder markers.

Table 6. Expression of *sdrP* and its target genes under various stress conditions

Gene name*	Spearman's correlation coefficient	Diamide	H_2O_2	$ZnSO_4$	Tetracycline	NaCl	Ethanol	$CuSO_4$ (wild type)	$CuSO_4$ (Δ*csoR*)	Stationary† phase
sdrP	1.00	27	10	8.1	6.1	5.6	5.2	2.1	18	22
TTHA1128	0.91	10	1.3	3.0	1.9	2.5	2.2	1.3	7.9	5.7
TTHA0986	0.83	12	2.5	2.7	0.8	1.7	1.1	1.2	7.8	7.6
TTHA0570	0.81	2.9	1.2	2.1	1.5	1.4	0.9	1.1	4.5	3.0
TTHA0029	0.80	6.2	1.6	2.6	1.3	1.7	2.5	1.2	5.5	2.8
TTHA1635	0.77	4.1	1.4	1.9	5.7	1.2	1.8	1.2	3.0	8.0
TTHA0637	0.77	3.7	1.4	1.6	0.6	1.3	1.0	1.1	2.7	2.3
TTHA0636	0.75	4.2	1.6	1.8	0.7	1.3	1.1	1.0	2.9	1.9
TTHA0557	0.72	3.5	1.5	1.6	2.2	1.3	1.0	0.9	3.7	2.9
TTHA1215	0.71	2.5	1.9	1.8	0.7	1.6	1.4	1.2	2.1	2.0
TTHA1892	0.70	3.5	3.0	2.2	0.8	0.9	1.8	1.1	2.4	3.2
TTHB132	0.67	2.6	1.8	2.0	1.4	0.8	1.3	1.1	1.7	4.0
TTHA1625	0.67	2.7	1.7	1.5	2.2	0.7	1.0	0.9	3.2	4.4
TTHA0654	0.67	3.9	1.7	1.7	0.3	1.6	0.9	1.2	3.4	2.0
TTHA0634	0.66	4.8	1.6	1.8	0.6	1.1	1.0	1.0	3.0	1.9
TTHA0635	0.66	4.0	1.6	1.6	0.6	1.1	1.2	1.0	3.3	1.7
TTHA0638	0.66	2.1	1.9	1.2	0.8	1.5	1.0	1.0	1.9	2.6
TTHA0769	0.65	2.0	2.3	1.2	0.4	1.3	1.6	0.8	2.0	2.8
TTHA0770	0.65	5.2	1.9	1.6	0.4	1.2	1.2	0.9	3.5	4.6
TTHA0337	0.51	1.9	2.0	0.6	1.2	1.1	1.0	1.0	2.8	1.5
TTHA0655	0.51	2.6	1.5	1.7	0.5	1.5	0.8	1.1	2.6	1.5
TTHA1028	0.48	1.7	1.3	0.7	1.4	1.4	0.7	1.0	2.6	1.4
TTHA0425	0.41	6.3	2.4	1.6	0.8	5.6	0.7	1.7	7.3	2.7

The expression level in cells after treatment with 2 mM diamide for 30 min, 10 mM H_2O_2 for 5 min, 1 mM $ZnSO_4$ for 30 min, 50 mM tetracycline for 10 min, 1.5% NaCl for 30 min, 5% ethanol for 30 min, or 1.25 mM $CuSO_4$ for 30 min, relative to untreated cells, is shown. The expression level in the stationary phase is the relative value compared to that in the logarithmic phase. Gray background indicates that the *q*-value is > 0.05. The details of these experiments are given in the NCBI GEO website under GEO series Accession Nos. described under the "RT-PCR analysis" subsection of the "Materials and methods" section. For the stationary phase data, that is GSE21290.

*The nucleotide sequences and deduced amino acid sequences can be found in the NCBI website accessible under GenBank Accession No. NC_006461, NC_006462, and NC_006463. The genes identified in this study are indicated by bold letters.

† The expression level after cultivation for 680 min relative to that for 300 min is shown, except in the case of the *TTHA0425* gene, the expression level of which is relative to that for 480 min (Agari *et al.*, 2008a).

Screening of SdrP-regulated genes by means of extensive expression pattern analysis

I found that expression of *sdrP* drastically changed depending on the environmental conditions. Because SdrP does not require any added effector molecule to induce transcription of its target genes *in vitro* (Fig. 12A and B) (Agari *et al.*, 2008a), the cellular responses via SdrP most likely depend on the amount of the SdrP, and not post-translational modification of the protein. In order to find additional SdrP-regulated genes, I performed expression pattern analysis using the 306 DNA microarray datasets derived with 117 experimental conditions, which were obtained for time-dependent expression analysis of the wild-type strain in a rich or synthetic medium (91 samples with 40 experimental conditions), expression analysis of a gene-disrupted strain (95 samples with 35 experimental conditions), expression analysis after chemical or physical treatment, or phage infection (87 samples with 29 experimental conditions), and a combination of gene-disruption with chemical or physical treatment, or with phage infection (33 samples with 13 experimental conditions) (Table 2). As a result, 40 genes whose expression was strongly positively correlated with that of the *sdrP* gene were selected, their Spearman's correlation coefficients being ≥ 0.65 (Fig. 17 and Table 8). Among them, the proportion of genes belonging to COGs (clusters of orthologous groups of proteins) code O (posttranslational modification, protein turnover, chaperones) and code C (energy production and conversion) were higher (Tables 7 and 8). Ten of the 14 SdrP-regulated genes identified by differential expression analysis of Δ*sdrP* strain (Table 5) were included in these 40 genes

(Tables 6 and 8).

On the other hand, expression of 16 genes was strongly and negatively correlated with that of the *sdrP* gene, with Spearman's correlation coefficients ≤ -0.65 (Tables 7 and 9). Among them, the proportion of genes belonging to COGs code H (coenzyme transport and metabolism) was the highest, suggesting that some specific metabolism was inversely correlated with the stress response via SdrP.

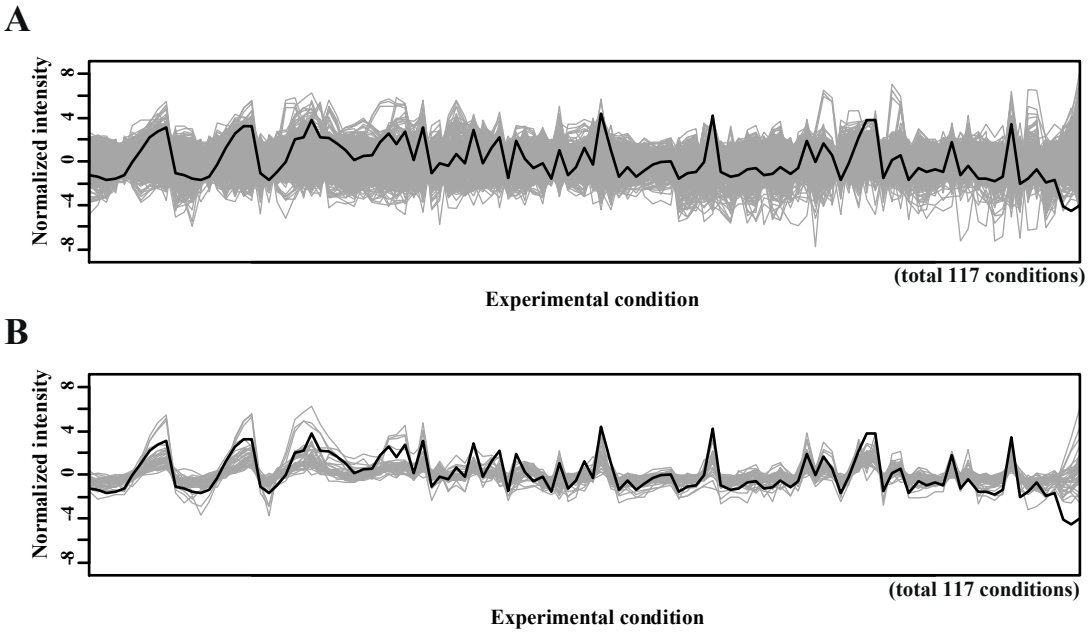


Fig. 17. Expression pattern analysis using DNA microarray data. The normalized intensity of each gene with each of the 117 experimental conditions is plotted. **(A)** All genes. **(B)** Genes whose Spearman's correlation coefficients were ≥ 0.65. Black line, *sdrP* gene; gray lines, the other genes. These graphs were generated using R (http://www.r-project.org).

Table 7. Number of genes correlated with *sdrP* gene in each COGs category

COGs code	Description	In genome*	Number of genes (% in total)							
			Spearman's correlation coefficient ≤ 1.0				Spearman's correlation coefficient ≥ -1.0			
			0.65 ≤	0.60 ≤	0.55 ≤	0.50 ≤	-0.65 ≥	-0.60 ≥	-0.55 ≥	-0.50 ≥
Information storage and processing										
J	Translation	153 (6.1)	-	-	2 (1.9)	4 (2.3)	1 (6.3)	1 (2.0)	6 (5.5)	9 (5.0)
K	Transcription	104 (4.1)	-	1 (1.6)	1 (0.9)	5 (2.9)	-	2 (3.9)	7 (6.4)	8 (4.4)
L	Replication, recombination and repair	115 (4.6)	1 (2.4)	1 (1.6)	2 (1.9)	4 (2.3)	-	2 (3.9)	6 (5.5)	9 (5.0)
B	Chromatin structure and dynamics	2 (0.1)	-	-	-	-	-	-	-	1 (0.6)
Cellular processes										
D	Cell cycle control, mitosis and meiosis	28 (1.1)	1 (2.4)	1 (1.6)	1 (0.9)	1 (0.6)	-	1 (2.0)	2 (1.8)	4 (2.2)
V	Defense mechanisms	20 (0.8)	-	-	1 (0.9)	3 (1.8)	-	-	-	1 (0.6)
T	Signal transduction mechanisms	65 (2.6)	2 (4.8)	2 (3.1)	2 (1.9)	3 (1.8)	-	1 (2.0)	1 (0.9)	1 (0.6)
M	Cell wall/membrane biogenesis	78 (3.1)	-	-	-	1 (0.6)	1 (6.3)	4 (7.8)	5 (4.6)	10 (5.6)
N	Cell motility	12 (0.5)	-	-	-	-	-	-	-	1 (0.6)
Z	Cytoskeleton	1 (0.0)	-	-	-	-	-	-	-	-
U	Intracellular trafficking and secretion	30 (1.2)	-	-	-	-	-	-	-	4 (2.2)
O	Posttranslational modification, protein turnover, chaperones	84 (3.3)	8 (19)	9 (14)	9 (8.4)	17 (9.9)	-	2 (3.9)	3 (2.8)	5 (2.8)
Metabolism										
C	Energy production and conversion	150 (6.0)	6 (14)	9 (14)	16 (15)	21 (12)	1 (6.3)	1 (2.0)	6 (5.5)	10 (5.6)
G	Carbohydrate transport and metabolism	125 (5.0)	2 (4.8)	6 (9.4)	8 (7.5)	14 (8.2)	1 (6.3)	2 (3.9)	3 (2.8)	8 (4.4)
E	Amino acid transport and metabolism	209 (8.3)	-	1 (1.6)	4 (3.7)	8 (4.7)	2 (12)	4 (7.8)	9 (8.3)	18 (10)
F	Nucleotide transport and metabolism	64 (2.6)	-	-	-	-	-	1 (2.0)	8 (7.3)	13 (7.2)
H	Coenzyme transport and metabolism	110 (4.4)	3 (7.1)	3 (4.7)	5 (4.7)	5 (2.9)	5 (31)	8 (16)	11 (10)	16 (8.9)
I	Lipid transport and metabolism	89 (3.5)	1 (2.4)	2 (3.1)	2 (1.9)	3 (1.8)	-	4 (7.8)	7 (6.4)	12 (6.7)
P	Inorganic ion transport and metabolism	91 (3.6)	1 (2.4)	3 (4.7)	4 (3.7)	7 (4.1)	1 (6.3)	2 (3.9)	4 (3.7)	6 (3.3)
Q	Secondary metabolite biosynthesis, transport and catabolism	58 (2.3)	-	1 (1.6)	6 (5.6)	7 (4.1)	-	-	1 (0.9)	2 (1.1)
Poorly characterized										
R	General function prediction only	305 (12)	4 (9.5)	6 (9.4)	11 (10)	18 (11)	3 (19)	7 (14)	10 (9.2)	15 (8.3)
S	Function unknown	166 (6.6)	5 (12)	6 (9.4)	10 (9.3)	12 (7.0)	-	5 (9.8)	9 (8.3)	11 (6.1)
-	Not in COGs	450 (18)	8 (19)	13 (20)	23 (21)	38 (22)	1 (6.3)	4 (7.8)	11 (10)	16 (8.9)
	Total	2509	42	64	107	171	16	51	109	180

*Number of genes in the genome of *T. thermophilus* HB8. The values in parentheses are the percentages of the total number of the genes.

Table 8. Genes exhibiting Spearman's correlation coefficients ≥ 0.65, as to the *sdrP* gene.

Gene name*	Spearman's correlation coefficient	Annotation for the product	COGs code†
TTHA1128	0.91	peptidase	R
TTHB088	0.85	Zn-dependent hydrolase	R
TTHA0986	0.83	highly conserved protein containing a thioredoxin domain	O
TTHA0164	0.83	thiol:disulfide interchange protein	OC
TTHA1804	0.82	acyl-CoA thioesterase	I
TTHA0570	0.81	glucose/sorbosone dehydrogenase	G
TTHA0029	0.80	hypothetical protein	-
TTHA0637	0.77	uncharacterized protein with a von Willebrand factor type A domain	R
TTHA0843	0.77	serine protein kinase	T
TTHA1635	0.77	iron-sulfur cluster biosynthesis protein IscA	S
TTHA0516	0.76	hypothetical protein	-
TTHA0841	0.76	stage V sporulation protein R (SpoVR)-related protein	S
TTHA1936	0.76	glycerol-3-phosphate ABC transporter, periplasmic glycerol-3-phosphate-binding protein	G
TTHA1803	0.76	pterin-4-alpha-carbinolamine dehydratase	H
TTHA1360	0.75	hypothetical protein	S
TTHA0636	0.75	nucleotidyltransferase substrate-binding protein-like protein	-
TTHA0665	0.75	N^5,N^{10}-methylenetetrahydromethanopterin reductase	C
TTHA0662	0.74	hypothetical protein	OC
TTHY7093	0.73	hypothetical protein	-
TTHA0557	0.72	superoxide dismutase [Mn]	P
TTHA1215	0.71	thioredoxin reductase	O
TTHB128	0.70	arsenite oxidase, small subunit	C
TTHA1892	0.70	excinuclease ABC subunit B (UvrB)	L
TTHB127	0.69	arsenite oxidase, large subunit	C
TTHA1492	0.69	cell division protein FtsH	O
TTHA1712	0.68	hypothetical protein	-
TTHB243	0.68	hypothetical protein	S
TTHA0434	0.68	hypothetical protein	-
TTHB132	0.67	methionine sulfoxide reductase A	O
TTHA1625	0.67	osmotically inducible protein OsmC	O
TTHA0654	0.67	MRP (multiple resistance and pH adaptation)-like protein	D
TTHA0635	0.66	predicted nucleotidyltransferase	R
TTHA0151	0.66	molybdopterin-converting factor, subunit 1 (MoaD)	H
TTHA0638	0.66	hypothetical protein	-
TTHA0520	0.66	NAD-dependent malic enzyme (malate dehydrogenase)	C
TTHA0842	0.66	hypothetical protein	S
TTHA0634	0.66	magnesium chelatase subunit Chl I	H
TTHA0770	0.65	ATP-dependent Lon protease	O
TTHA0769	0.65	DegQ, trypsin-like serine protease	O
TTHA0419	0.65	hypothetical protein	R

*The nucleotide sequences and deduced amino acid sequences can be found in the NCBI website under GenBank Accession No. NC_006461, NC_006462, and NC_006463. The genes designated as TTHY are not present in the genomic analysis results.

†J, translation; K, transcription; L, replication, recombination and repair; B, chromatin structure and dynamics; D, cell cycle control, mitosis and meiosis; V, defense mechanisms; T, signal transduction mechanisms; M, cell wall/membrane biogenesis; N, cell motility; Z, cytoskeleton; U, intracellular trafficking and secretion; O, posttranslational modification, protein turnover, chaperones; C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; Q, secondary metabolite biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown; -, not in COGs.

Table 9. Genes exhibiting Spearman's correlation coefficients ≤ -0.65, as to the *sdrP* gene.

Gene name*	Spearman's correlation coefficient	Annotation for the product	COGs code†
TTHA0319	-0.74	hypothetical protein	R
TTHA0098	-0.73	arginyl-tRNA synthetase	J
TTHA1438	-0.70	hypothetical protein	HR
TTHB048	-0.69	nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferase	H
TTHA0424	-0.69	thiamine-monophosphate kinase	H
TTHA0457	-0.68	3-phosphoshikimate 1-carboxyvinyltransferase	E
TTHA1304	-0.68	sugar ABC transporter, ATP-binding protein	R
TTHA1282	-0.67	hypothetical protein	-
TTHA1120	-0.67	methylenetetrahydrofolate dehydrogenase (NADP+)/methenyltetrahydrofolate cyclohydrolase	H
TTHA1115	-0.67	integral membrane protein, TerC family	P
TTHA0879	-0.67	putative oligosaccharide deacetylase	G
TTHA1314	-0.67	UDP-N-acetylglucosamine 2-epimerase	M
TTHA0191	-0.67	dihydropteroate synthase	H
TTHA1275	-0.66	V-type ATP synthase subunit	C
TTHA1152	-0.65	aminopeptidase T	E
TTHA1574	-0.65	putative dehydrogenase	HE

*The nucleotide sequences and deduced amino acid sequences can be found in the NCBI website under GenBank Accession No. NC_006461, NC_006462, and NC_006463.

†J, translation; K, transcription; L, replication, recombination and repair; B, chromatin structure and dynamics; D, cell cycle control, mitosis and meiosis; V, defense mechanisms; T, signal transduction mechanisms; M, cell wall/membrane biogenesis; N, cell motility; Z, cytoskeleton; U, intracellular trafficking and secretion; O, posttranslational modification, protein turnover, chaperones; C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; Q, secondary metabolite biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown; -, not in COGs.

Identification of novel SdrP regulated genes

In order to determine whether novel SdrP-regulated genes are included in the 40 genes that showed Spearman's correlation coefficients of ≥ 0.65, I searched for SdrP-binding sites upstream of these genes. I found that sequences upstream of the *TTHA0029*, *TTHA0557*, *TTHA1128*, *TTHA1215*, *TTHA1625*, *TTHA1635*, *TTHA1892*, and *TTHB132* genes were homologous to that of a putative consensus SdrP-binding site (Fig. 15B). The DNA fragments containing the putative binding sites were cloned and used as templates for *in vitro* run-off transcription assays. I found that all of the genes were transcribed by *T. thermophilus* RNAP-σ^A holoenzyme in an SdrP-dependent manner, as in the cases of the differential expression analysis (Fig. 18). SdrP did not enhance transcription of the DNA fragment containing upstream of the *TTHA0987* gene (Spearman's correlation coefficient = 0.64) (Fig. 18), or those containing other genes derived from *T. thermophilus* HB8 (Agari *et al.*, 2008a), indicating that SdrP does not nonspecifically bind DNA under these experimental conditions. I found that the *in vitro* transcription start sites of the novel SdrP-regulated genes were 6–7 bp downstream from the predicted -10 hexamers of their promoters and around 40 bp downstream of the putative SdrP-binding sites (Figs 14B and 15B). I investigated the sequence conservation of the putative binding-sequences of 16 SdrP-regulated promoters including those identified in the differential expression analysis (Fig. 15). The results indicate that the left arm of the putative binding-sites is relatively conserved as TTGTG, but the right arm is not except

for two C bases (Fig. 15C).

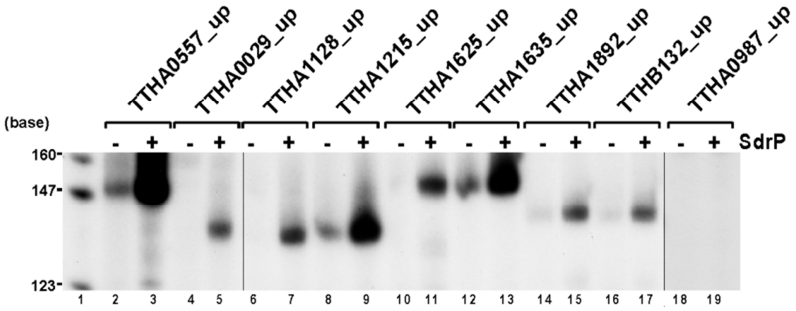


Fig. 18. Run-off transcription assays performed with a template containing the upstream sequence of the *TTHA0557* (TTHA0557_up), *TTHA0029* (TTHA0029_up), *TTHA1128* (TTHA1128_up), *TTHA1215* (TTHA1215_up), *TTHA1625* (TTHA1625_up), *TTHA1635* (TTHA1635_up), *TTHA1892* (TTHA1892_up), *TTHB132* (TTHB132_up), or *TTHA0987* (TTHA1987_up) gene in the absence (-) or presence (+) of *T. thermophilus* SdrP. After the reaction, equivalent volumes of samples were analyzed by PAGE followed by autoradiography. Lane 1, [α-^{32}P]-dCTP-labelled MspI fragments of pBR322.

Table 10 summarizes the eight genes that are under the control of the SdrP-dependent promoter found in this analysis. Expression of the genes also tended to increase upon entry into the stationary phase, as in the case of the previously identified SdrP-regulated genes (Table 6). I could not find the predicted SdrP-binding sequence close to the promoter regions of the 16 genes whose expression showed strong negative correlation with that of the *sdrP* gene, suggesting that SdrP does not act as a transcription repressor. Thus, including the 14 previously identified genes, a total of 22 genes have been identified as SdrP-regulated genes (Table 11).

I analyzed the altered expression profiles of the 22 SdrP-regulated genes in cells perturbed by the various stresses, and found that the expression of most genes increased with these perturbations (Table 6). The altered expression profile caused by 2 mM diamide treatment was the most similar to that upon entry into the stationary phase (Table 6). The expression level

did not always correlate with that of the *sdrP* gene, especially in response to perturbation by 50 mM tetracycline, in which the expression of 13 genes was significantly decreased (Table 6). These results suggest that depending on the stress, not only the signal via SdrP, but also other signal(s) are transmitted to the cells to alter expression of the SdrP-regulated genes.

Table 10. Genes under the control of the SdrP-dependent promoters identified by means of expression pattern analysis

Gene name	Conserved domain*	*e*-value†	Annotation of product	Possible cellular role‡	Reference
TTHA0029	-	-	Hypothetical protein	Transcriptional regulation	
TTHA0557	COG0605	8e-73	Manganese superoxide dismutase	Redox control	(Ludwig *et al.*, 1991; Peterson *et al.*, 1991)
TTHA1128	COG1506	2e-21	Peptidase	Protein quality control; Nutrient supply	
TTHA1215	COG0492	2e-30	Thioredoxin reductase	Redox control	
TTHA1625	TIGR03562	7e-54	Catalase	Redox control	(Rehse *et al.*, 2004)
TTHA1635	TIGR00049	8e-41	Iron-sulfur cluster biosynthesis protein IscA	Redox control	(Yang *et al.*, 2006; Zheng *et al.*, 1998)
TTHA1892	COG0556	1e-180	Excinuclease ABC subunit B (UvrB)	DNA repair	(Nakagawa *et al.*, 1999)
TTHB132	PRK00058	2e-72	Methionine sulfoxide reductase A (MsrA)	Protein repair	(Ezraty *et al.*, 2005)

*The domain was identified in a BLAST search.
†This value is for the amino acid sequence of the conserved domain.
‡The C-terminal domain of the TTHA0029 protein showed homology to that of *Meiothermus silvanus* ZP_04036762 annotated as a transcriptional regulatory protein, the *e*-value being 3e-20.

1-5. Discussion

Thermus thermophilus SdrP is one of four CRP/FNR family proteins from the extremely thermophilic bacterium *T. thermophilus* HB8, and its expression increases in the stationary phase during cultivation at 70°C. A BLAST search revealed that bacteria from the phylum *Deinococcus-Thermus* had many proteins that showed high similarity to SdrP. In the case of *T. thermophilus* CRP, which is a cAMP-dependent transcriptional activator and exhibits 43% similarity (*e*-value = 1e-9) to SdrP, many proteins with the highest levels of similarity were from the phyla *Firmicutes* and *Cyanobacteria*, not the phylum *Deinococcus-Thermus* (Shinkai *et al.*, 2007a). These results suggest that SdrP is evolutionarily different from CRP.

The cAMP-dependent regulatory mechanism of CRP has been extensively studied for *E. coli* CRP, a prototype of this family of proteins (Busby & Ebright, 1999; Kolb *et al.*, 1993; Lawson *et al.*, 2004). *E. coli CRP* is a homodimer that contains a helix-turn-helix DNA-binding motif in its C-terminal domain. CRP undergoes conformational change upon cAMP binding, and the CRP-cAMP complex interacts with a 22 bp DNA site exhibiting twofold symmetry that has the consensus sequence 5′-AAATGTGATCTAGATCACATTT-3′ (Ebright *et al.*, 1989). The crystal structure of *T. thermophilus* SdrP is similar to that of the DNA-binding form of *E. coli* CRP – the form complexed with cAMP and DNA – and has an r.m.s.d. of 2.3 Å. In addition, the structure of SdrP suggests that cAMP cannot enter the site corresponding to the primary cAMP binding site of *E. coli* CRP due to steric hindrance by bulky residues. These structural

properties of SdrP imply that this protein does not require an effector molecule to bind DNA, which is supported by the observation that this protein can positively regulate transcription independent of any effector molecule *in vitro*. Y70 of SdrP, which probably causes steric hindrance in cAMP binding, corresponds to S84 of *E. coli* CRP and possibly S86 of *T. thermophilus* CRP (Figs 8 and 10C). At this position, a small residue might be necessary for a CRP family protein to act as a cAMP-dependent transcriptional regulator.

Escherichia coli YeiL is a CRP/FNR family protein, and its expression increases in the stationary phase (Anjum *et al.*, 2000). YeiL has been suggested to have an iron-sulfur center and a reversible intra- and interchain disulfide bond (Anjum *et al.*, 2000), while SdrP does not have cysteine residues. Expression of the *yeiL* gene is dependent on an alternative σ factor, i.e. σ^S, and is positively autoregulated and influenced by FNR (Anjum *et al.*, 2000). I found that a housekeeping σ, i.e., σ^A, but not a sole alternative σ, i.e., σ^E, was involved in the transcription of the SdrP regulon and that its expression was not autoregulated. Therefore, the regulatory mechanism involving SdrP may differ from that for YeiL.

The crystal structure of the *E. coli* CRP-DNA complex revealed that one subunit of the CRP dimer binds to the left arm of the ^{4}TGTGA block, while the other binds to the right arm of the ^{15}TCACA block of the 22 bp consensus CRP binding site. The R181, E182, and R186 residues of *E. coli* CRP directly interact with the G:C pairs at positions 5 and 7 and with the A:T pair at position 8 of the consensus CRP binding half site (Parkinson *et al.*, 1996; Schultz *et al.*, 1991).

In *E. coli*, based on the position of the CRP binding site relative to the transcriptional start site, simple CRP-dependent promoters, which require one CRP dimer for transcriptional activation, are grouped into two classes, i.e., I (position -61.5) and II (position -41.5) (Busby & Ebright, 1999; Lawson *et al.*, 2004). Based on these properties of *E. coli* CRP and the fact that the crystal structure of SdrP is similar to that of the DNA-binding form of *E. coli* CRP (Fig. 10), I predicted SdrP-binding sequences [consensus sequence: 5′-TTGTG(N7-9)CxCxx-3′] in the SdrP- dependent promoters (Fig. 15C), which are similar to the consensus binding sites for *E. coli* CRP and other characterized members of the CRP/FNR family (Bai *et al.*, 2005; Cameron & Redfield, 2006; Ebright *et al.*, 1989; Hsiao *et al.*, 2005; Kanack *et al.*, 2006; Letek *et al.*, 2006). The binding site should be defined by biochemical experiments (such as the footprinting assay); however, if this prediction is correct, the SdrP-dependent promoters are more similar to class II than to class I CRP-dependent promoters of *E. coli* (Fig. 15C).

Of the genes that exhibited lower expression in the *T. thermophilus* Δ*sdrP* strain, I identified eight SdrP-dependent promoters that were upstream of the *TTHA0337*, *TTHA0425*, *TTHA0570*, *TTHA0770*, *TTHA0634*, *TTHA0654*, *TTHA0986*, or *TTHA1028* gene *in vitro*. I observed that the expression of the eight genes tended to increase in the stationary phase compared with the logarithmic phase during cultivation in a rich medium, although their mRNA expression profiles did not necessarily parallel that of *sdrP* mRNA. In addition to SdrP, some other regulatory factors might be involved in the expression of these genes.

Expression of genes downstream of *TTHA0634*, *TTHA0654*, and *TTHA0770* was also decreased in the Δ*sdrP* strain (Table 5). In the wild-type strain, their expression tended to increase in the stationary phase during cultivation in a rich medium: expression in the stationary phase (680 min cultivation) was 1.5- to 2.8-fold higher than that in the logarithmic phase (300 min cultivation) (Table 6). Therefore, such genes may form operons such as *TTHA0634-TTHA0635-TTHA0636-TTHA0637-TTHA0638*, *TTHA0654-TTHA0655* and *TTHA0770-TTHA0769*, as indicated by genome analysis (GenBank Accession No. NC_006461).

Furthermore, of the genes whose expression were strongly positively correlated with that of the *sdrP* gene, I identified additional eight SdrP-dependent promoters that were upstream of the *TTHA0029*, *TTHA0557*, *TTHA1128*, *TTHA1215*, *TTHA1625*, *TTHA1635*, *TTHA1892*, or *TTHB132* gene *in vitro*. These eight genes were not identified as SdrP-regulated genes in differential expression analysis for the following reasons. Although the expression levels of the eight genes were 0.17–0.63-fold in the Δ*sdrP* strain relative to that in wild type, their *q*-values except that of *TTHA1128* were 0.061–0.242, which were greater than the threshold value used in the experiment (0.06). As for *TTHA1128*, identification of an SdrP-binding site in the promoter region was missed in the analysis. Conversely, expression of four out of 14 SdrP-regulated genes identified in the differential expression analysis of Δ*sdrP* strain showed lower correlation to that of *sdrP* (Spearman's correlation coefficients ≤ 0.51) (Table 6). Some unknown factors such as promoter activity and affinity of SdrP to DNA *in vivo*, and unidentified

transcriptional regulator(s) that might act together with SdrP, might influence the results of the experimental screenings for SdrP-regulated genes. Thus, a combination of comparative expression analysis and expression pattern analysis was appropriate for screening of SdrP-regulated genes.

To determine the cellular role of SdrP and the stationary-phase physiology of the *T. thermophilus* HB8 strain, I predicted the molecular functions of the SdrP-regulated gene products by investigating their amino acid sequences and structural features because many of these are biochemically or biophysically uncharacterized (Table 11). The possible cellular roles of the SdrP-regulated gene products, with the exception of five functional unknown proteins, are roughly classified into four groups: energy or nutrient supply, polyadenylation of mRNA, repair and/or turnover of DNA and protein, and redox control (Table 11). I discussed the cellular roles of SdrP and its target gene products below. The *T. thermophilus* Δ*sdrP* strain exhibited a growth defect, especially in a synthetic medium (Fig. 11A and B). With a limited amount of nutrients, i.e., in a synthetic medium or stationary phase, insufficient nutrients and energy might be supplied through the activity of some SdrP-regulated gene products. It has been shown that polyadenylation of mRNA is possibly an important factor that promotes adaptation to slow growth conditions such as those under which the amount of nutrients is limited (Jasiecki & Wegrzyn, 2003; Santos *et al.*, 2006). Oxidative damage may occur to DNA and protein in *T. thermophilus*, especially in the stationary phase, as observed in many bacteria

(Ballesteros *et al.*, 2001; Bridges, 1993; Nyström, 2004). One major effect of such oxidative damage is the oxidation of thiols that results in non-native disulfide bond formation in proteins. I observed that the growth sensitivity of *T. thermophilus* HB8 under disulfide stress conditions increased when the *sdrP* gene was deleted (Fig. 11B). The enzymes involved in redox control, such as manganese superoxide dismutase (Ludwig *et al.*, 1991; Peterson *et al.*, 1991), catalase (Rehse *et al.*, 2004), and thioredoxin reductase, may protect cellular components from oxidative damage. The excinuclease ABC subunit B (UvrB) plays a central role in the nucleotide excision repair of damaged DNA (Nakagawa *et al.*, 1999). The methionine sulfoxide reductase A (MsrA) acts to reduce the methionine sulfoxide generated by the oxidation of a methionine residue, which is the most sensitive amino acid residue to reactive oxygen species (Ezraty *et al.*, 2005). The proteases and peptidase may be involved in the turnover of damaged proteins. According to the BLAST searches, SdrP-regulates two functionally unknown proteins, TTHA0655 and TTHA0029, which were predicted to be transcriptional regulators (Table 11). These proteins might control some of the genes with expression that is positively or negatively correlated with that of *sdrP*, but not directly regulated by SdrP (Table 4). The three remaining functionally unknown proteins in Table 11 might participate in nutrient supply, mRNA polyadenylation, redox control or repair/turnover of DNA or proteins, because in many cases, the cellular functions of the genes regulated by a certain transcriptional factor are similar.

Table 11. Summary of the genes under the control of the SdrP-dependent promoters

Gene name	Annotation for product	Possible cellular role
TTHA0425	NAD(P)H oxidase	energy/nutrient supply; redox control
TTHA0570	glucose/sorbosone dehydrogenase	energy/nutrient supply
TTHA0654	MRP (multiple resistance and pH adaptation)-like protein	energy/nutrient supply; redox control
TTHA0634	magnesium chelatase subunit Chl I	polyadenylation of mRNA (chelation of metal ion to TTHA0635/0636)
TTHA0635	predicted nucleotidyltransferase	polyadenylation of mRNA
TTHA0636	nucleotidyltransferase substrate-binding protein-like protein	polyadenylation of mRNA
TTHA0557	manganese superoxide dismutase	redox control
TTHA0986	highly conserved protein containing a thioredoxin domain	redox control
TTHA1028	SseA, rhodanese-related sulfurtransferase	redox control
TTHA1215	thioredoxin reductase	redox control
TTHA1625	catalase	redox control
TTHA1635	iron-sulfur cluster biosynthesis protein IscA	redox control
TTHA0769	DegQ, trypsin-like serine protease	protein repair/turnover
TTHA0770	ATP-dependent Lon protease	protein repair/turnover
TTHA1128	peptidase	protein repair/turnover
TTHB132	methionine sulfoxide reductase A (MsrA)	protein repair/turnover, redox control
TTHA1892	excinuclease ABC subunit B (UvrB)	DNA repair/turnover
TTHA0029	hypothetical protein	unknown
TTHA0655	predicted transcriptional regulator	unknown
TTHA0337	hypothetical protein	unknown
TTHA0637	uncharacterized protein with a von Willebrand factor type A domain	unknown
TTHA0638	hypothetical protein	unknown

Among the environmental and chemical stresses examined in this study, the diamide and H_2O_2 stresses were the most effective in enhancing the expression of *sdrP* and its target genes in the wild-type strain (Table 6). Furthermore, an excess amount of $CuSO_4$ was a strong inducer of *sdrP* gene expression in the Δ*csoR* strain, in which excess Cu(I) ions may accumulate (Sakamoto *et al.*, 2010). In this strain, excess Cu(I) ions, which have the potential to drive

oxidation/reduction to form free radicals (Imlay, 2002; Touati, 2000), may trigger expression of *sdrP*. As for the possible cellular functions of the 22 SdrP-regulated gene products, at least nine proteins are possibly involved in redox control; five proteins may be involved in the repair or turnover of oxidized cellular component (Table 11). The altered expression levels of *sdrP* and its target genes in the stationary phase were similar to those caused by diamide treatment (Table 6). These results suggest that the main inducer of *sdrP* expression is oxidative stress, and support the finding that SdrP functions in the response to oxidative stress. Because SdrP does not have a cysteine residue or cofactor that could be a sensor of an oxidative signal [unlike in the case of other oxidative stress-responsive transcriptional regulators such as OxyR, PerR, and SoxR (Lee & Helmann, 2006; Pomposiello & Demple, 2001; Storz & Imlay, 1999)], and it does not require any effector molecule for its transcriptional activation (Figs 12A and B, and 18), there may be some unidentified factor(s) sensing oxidative stress and causing induced expression of SdrP. Elucidation of the regulatory signal that induces SdrP expression in the stationary phase will facilitate understanding of fundamental stress response physiology of *T. thermophilus* as well as the physiological function of SdrP.

Chapter 2

Stress response in the phage infection

2-1 Abstract

The clustered regularly interspaced short palindromic repeat (CRISPR) systems composed of DNA direct repeats designated as CRISPRs and several CRISPR-associated (*cas*) genes, which are present in many prokaryotic genomes, comprise a host defense system against invading foreign replicons such as phages. In order to investigate the altered expression profiles of the systems after phage infection using a model organism, *Thermus thermophilus* HB8, which has 12 CRISPR loci, genome-wide transcription profiling of the strain infected with lytic phage *Φ*YS40 was performed by DNA microarray analysis. Significant alteration of overall mRNA expression gradually increased during infection, i.e., from the eclipse period to the period of host cell lysis. Interestingly, the expression of most cAMP receptor protein (CRP)-regulated genes, including two CRISPR-associated (*cas*) operons, was most markedly up-regulated, especially around the beginning of host cell lysis, although up-regulation of the *crp* gene was not observed. Expression of the CRP-regulated genes was less up-regulated in a *crp*-deficient strain than in the wild type. Thus, it is suggested that cAMP is a signaling molecule that transmits information on phage infection to CRP to up-regulate these genes. On the other hand, expression of several *cas* genes and that of CRISPRs were up-regulated independent of CRP, suggesting the involvement of unidentified regulatory factor(s) induced by phage infection. On analysis of the expression profile of the entire genome, I could speculate that upon phage infection, the signal was transmitted to the cells, host response systems including CRISPR

defense systems being activated, while the overall efficiencies of transcription, translation, and metabolism in the cells decreased. These findings will facilitate understanding of the host response mechanism following phage infection.

2-2 Introduction

Bacteriophages have a major influence on the microbial world (Chibani-Chennoufi *et al.*, 2004). Phages regulate host macromolecular synthesis by modifying the transcription and translation machineries to propagate in the cells. In contrast, bacteria have developed various defense systems against phage infection. The clustered regularly interspaced short palindromic repeat (CRISPR) systems that are present in many prokaryotic genomes are a recently discovered host defense system (Barrangou *et al.*, 2007; Horvath & Barrangou, 2010; Makarova *et al.*, 2006; Marraffini & Sontheimer, 2010; Sorek *et al.*, 2008; van der Oost *et al.*, 2009). In general, these systems are composed of CRISPR and CRISPR-associated (*cas*) genes (Haft *et al.*, 2005; Jansen *et al.*, 2002). CRISPRs are composed of 24- to 47-bp direct repeats separated by nonrepetitive unique spacer sequences of similar length. Sequences derived from foreign replicons such as phages and plasmids are found in the spacers of several CRISPRs (Bolotin *et al.*, 2005; Mojica *et al.*, 2005; Pourcel *et al.*, 2005). CRISPR loci are transcribed and processed into small CRISPR RNAs (Brouns *et al.*, 2008; Hale *et al.*, 2008) that specify acquired immunity against foreign replicons through a mechanism that relies on the strict identity between CRISPR spacers and targets (Barrangou *et al.*, 2007; Brouns *et al.*, 2008; Hale *et al.*, 2008). If cells do not have any CRISPR spacers that are identical with the sequences of an invading replicon, a fragment derived from the replicon can be incorporated into a CRISPR locus of the cells as a new spacer after infection by the replicon; this phase is designated as the

adaptation phase. The new spacer sequence plays a role in immunity against subsequent infection by the same foreign replicon, this phase being designated as the interference phase. Bioinformatical and experimental studies support that transcribed spacer RNAs directly target DNA or RNA of foreign replicons (Makarova *et al.*, 2006; Marraffini & Sontheimer, 2008; Shah *et al.*, 2009; Sorek *et al.*, 2008). Several CRISPR system subtypes have been found, and Cas proteins are classified into about 45 families based on their amino acid sequences, most of which remain uncharacterized (Haft *et al.*, 2005). The Cas protein families are mainly categorized into core proteins, eight subtypes, or the repeat-associated mysterious protein (RAMP) modules (Haft *et al.*, 2005). Core *cas* genes *cas1* to *cas6* are localized in close proximity to CRISPRs, and they are widely distributed in bacterial and archaeal genomes. The amino acid sequences or three-dimensional structures of several Cas proteins are similar to those of RNA- or DNA-binding proteins. In fact, Cas1 is a metal-dependent DNase, and it has been thought to be involved in the acquisition of new CRISPR spacers (Wiedenheft *et al.*, 2009). Cas2 cleaves single-stranded RNAs preferentially within U-rich regions (Beloglazova *et al.*, 2008). Cas3 and Cas4 resemble helicase and the RecB family of exonucleases, respectively (Jansen *et al.*, 2002). Cas6 (Carte *et al.*, 2010) and the Ecoli subtype Cas protein complex, also called the cascade complex, which consists of Cse1, Cse2, Cse3, Cse4, and Cas5e (Brouns *et al.*, 2008), cleave a CRISPR RNA precursor in each repeat, with the cleavage products being retained. The X-ray crystal structure of Cse3 is similar to those of many RNA-binding

proteins (Ebihara *et al.*, 2006). Cse2 and a RAMP module, Cmr5, adopt a novel fold with large continuous basic patches on one side of their surfaces, which possibly bind DNA or RNA (Agari *et al.*, 2008b; Sakamoto *et al.*, 2009). Although the molecular mechanism of the CRISPR systems is beginning to be studied, details including those of transcriptional regulation of these systems remain unknown.

T. thermophilus HB8 has ten (CRISPR-1–CRISPR-10) and two (CRISPR-11 and CRISPR-12) CRISPR loci on pTT27 and the chromosome, respectively (Fig. 19); that is, four loci have been identified in addition to the eight reported previously (Godde & Bickerton, 2006). On pTT27, several core *cas* genes (i.e., *cas*1–*cas*4 and *cas*6), one set each of the Ecoli and Mtube subtypes, RAMP module genes, and several other *cas* family genes are encoded (Fig. 19) (Haft *et al.*, 2005). Of them, an operon containing *cas1* to *cas3* and Ecoli subtype genes and one containing Mtube subtype and *cas* family genes are positively regulated by cAMP receptor protein (CRP) in a cAMP-dependent manner (Fig. 19) (Shinkai *et al.*, 2007a). *Φ*YS40, a lytic tailed myophage that infects *T. thermophilus* HB8 (Sakaki & Oshima, 1975) (Fig. 20A), is the most characterized at the molecular level among 115 thermophages (Yu *et al.*, 2006). Its genome sequence has been determined (Naryshkina *et al.*, 2006), and regulation of gene expression was investigated by means of DNA macroarray analysis, with three temporal classes of phage genes (i.e., early, middle, and late) being identified (Sevostyanova *et al.*, 2007). The sequences of phage DNA are not found in the CRISPR spacer sequences of *T. thermophilus*

HB8.

Here, I show alteration of the transcription profile of the host cells during infection with *Φ*YS40 by DNA microarray analysis, focusing on the CRISPR systems. This study provides a novel insight into the host response mechanism following phage infection.

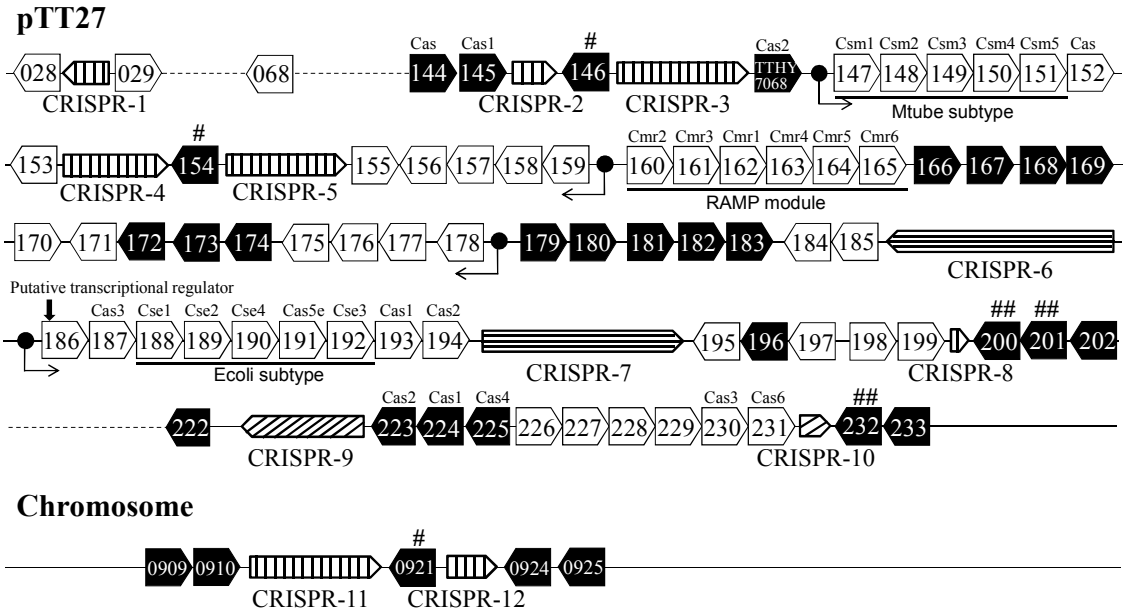


Fig. 19. Schematic representation around CRISPRs on pTT27 and the chromosome of *T. thermophilus* HB8. Expression increased genes at 100 min ($q \leq 0.05$, on ORF-level analysis) in the phage-infected *T. thermophilus* HB8 wild-type strain are denoted by white arrowheads, and the genes whose altered expression was not observed are shown by black arrowheads. The numerals represent '*TTHB*'- and '*TTHA*'-omitted gene names on pTT27 and the chromosome, respectively. *TTHY7068*, previously reported as *TT1823* (Beloglazova *et al.*, 2008), is not present in a new version of the genomic analysis results. *cas* gene names are shown above the arrowheads. Probable transposase genes are indicated by # or ##, those with the same marks having the same or similar sequences. Black circles with arrows indicate the positions of the CRP-dependent promoters and their transcriptional direction. CRISPRs are shown as patterned arrowheads, those with the same patterns having the same or similar repeat sequences (Table 13).

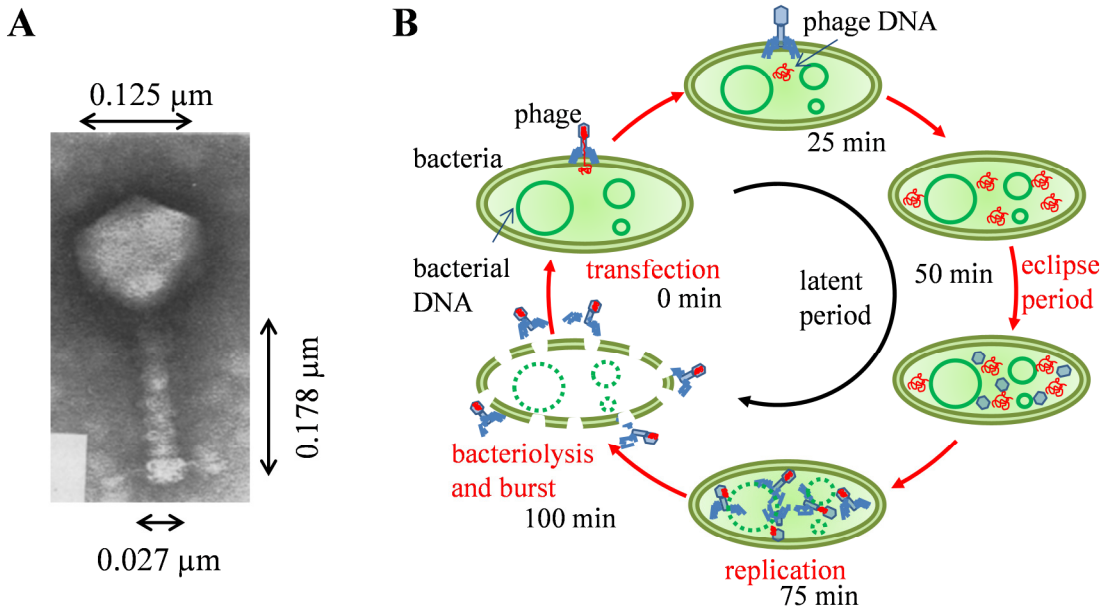


Fig. 20. (A) Electron micrographs of *Φ*YS40 negatively stained with uranyl acetate (Sakaki & Oshima, 1975). **(B)** Schematic diagram of the life cycle of *Φ*YS40 phage and the progeny phage production.

2-3. Materials and methods

Strains, cell growth and phage infection

The *crp*-deficient (Δ*crp*) *T. thermophilus* HB8 strain was constructed as described previously (Shinkai *et al.*, 2007a). The *Φ*YS40 phage stock solution (0.3–1.9 × 10^{11} pfu ml^{-1}) was prepared basically as described (Sevostyanova *et al.*, 2007) except that the strain was cultured in TT broth (see "Media and growth conditions for *T. thermophilus* HB8" subsection of section 1.3) at 70°C. *T. thermophilus* HB8 strains were cultured in 1 l of TT broth at 70°C until the A_{600} value reached ~0.8 (1.7 × 10^8 cells ml^{-1}), which corresponded to the logarithmic growth phase. For DNA microarray analysis, cells were collected from 50 ml of the culture medium, and then the *Φ*YS40 phage was introduced into the remaining medium at a multiplicity of infection of ~1, and the cultivation was continued. After 25, 50, 75, and 100 min, cells were collected from 50 ml of the culture medium (Sevostyanova *et al.*, 2007). For the premature lysis experiment, 1 ml of the broth at each time point was collected, and then a few drops of chloroform were added to the broth. After heating at 70°C for 10 min, a plaque assay was performed.

Improvement of the Chip Definition File of the *Thermus thermophilus* HB8 GeneChip

The genomic sequences in public databases, such as GenBank, EMBL-Bank, and DDBJ, are occasionally updated; for example, the protein-coding region, the replication origin, or the

number of genes is sometimes revised. Along with the revision, the DNA microarray design should be improved (Dai *et al.*, 2005; Lu & Zhang, 2006; Sandberg & Larsson, 2007). Since the genome sequence of *T. thermophilus* HB8 has also been updated, I revised the design of the *T. thermophilus* GeneChip (TTHB8401a520105F, Affymetrix). The *T. thermophilus* GeneChip has 25 mer single-stranded DNA probes corresponding to the nucleotide sequences of the transcripts (perfect match probes PM), and also has mismatched probes (MM) that contain a single base mismatch in the center of each probe, to detect non-specific signals (Fig. 21). The set of PM and MM is called a probe pair. This GeneChip carries 16 probe pairs for each ORF, which are called the probe sets (Fig. 21). The expression level of an ORF is estimated from the fluorescence intensities of the probe set. In addition, the probe set contains the probe pairs for protein non-coding (intergenic) regions. The genomic coordinates of each probe pair and the composition of the probe pairs in each probe set are defined by the Chip Definition File. I found that this GeneChip contained several probes with sequences that are not present in the current genomic sequence. In addition, the GeneChip contained redundant probes, as well as some that could hybridize with multiple loci. Therefore, I eliminated these useless (in total 4,698) probe pairs from the Chip Definition File. Furthermore, I re-organized the probe set definition with the remaining probes. The resulting new Chip Definition File was then formatted as "tthb8401a520105f.cdf Rev.2". The probe set definition in the new Chip Definition File has been deposited in the GEO, and is accessible through GEO Platform

Accession No. GPL9209.

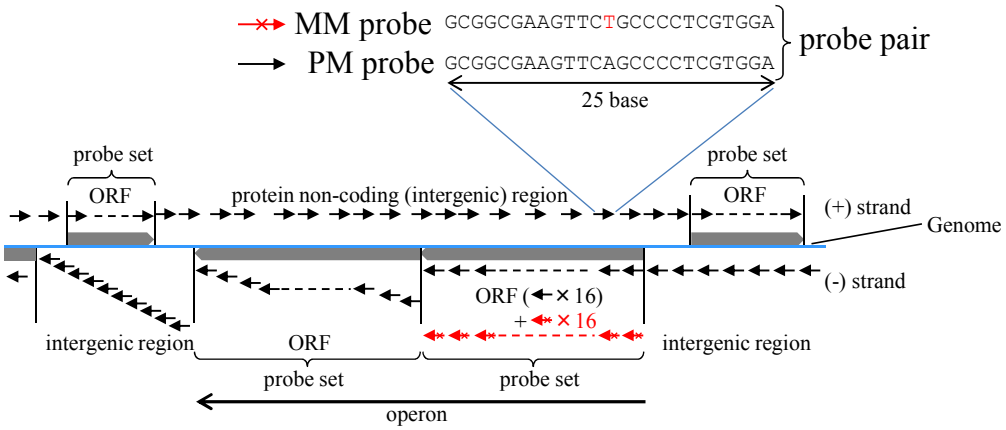


Fig. 21. Schematic diagram of the design of *T. thermophilus* GeneChip.

DNA microarray analysis

(i) *Sample preparation and data collection.* Sample preparation and data collection for the DNA microarray analysis were performed by the same procedures as described in Section 1-3 (i) of Chapter 1.

(ii) *Expression analysis at the ORF level.* The raw intensities for three independently cultured lots of non-infected and three lots of post-infected cells were each summarized as ORFs, using the GeneChip Operating Software, version 1.2 (Affymetrix). The datasets were then normalized through the following three normalization steps, using the GeneSpring GX 7.3.1 program (Agilent Technologies); i.e., data transformation (shifting of low signals < 0.01 to 0.01), followed by global scaling (normalization to the median of each array), and normalization using the data for the non-infected cells as a control sample, as described in Section 1-2 (iv) of

Chapter 1. I excluded several genes with detection calls that were 'Absent' (Pepper *et al.*, 2007) for all 15 wild-type and nine Δ*crp* strains. The remaining data for 2,202 and 2,181 ORFs of the wild-type and Δ*crp* strains, respectively, were used for the following analysis. The false discovery rate (*q*-value) (Storey, 2002) of the observed differences in the normalized intensities between the non- and post-infected cells was calculated using R (http://www.r-project.org).

(iii) *Expression analysis at the probe level.* In order to investigate the expression of intergenic regions, I established a method for the probe level expression analysis of the GeneChip data, as described below. The PM probe intensities of three sets of post-infected data at a particular time point and three sets of non-infected data were simultaneously quantile normalized (Bolstad *et al.*, 2003). In order to determine the expression differences between non-infected and post-infected cells, the mean values of the three probe intensities of the post-infected data were each divided by those of the non-infected ones. Then, the Wilcoxon signed-rank test was applied with a ± 100-bp window positioned at the center of each probe, which gave the *p*-value of the center probe (Cawley *et al.*, 2004). Data processing, statistical analysis, and data visualization were performed using R and a Bioconductor (Gentleman *et al.*, 2004). The microarray data presented in this study have been deposited in the GEO, and are accessible through GEO Series Accession No. GSE16978.

2-4. Results

Alteration of the overall mRNA expression profile after phage infection.

The overall mRNA expression in *T. thermophilus* HB8 infected by phage *Φ*YS40 at 25, 50, 75, and 100 min post-infection was analyzed at the ORF level and compared with that in non-infected cells. I confirmed that 99.8% of the phage population had infected the cells at 25 min (data not shown), the phage was still in the eclipse period at 50 min, phage progeny began to be produced at around 75 min, and host cell lysis began at around 100 min (Fig. 22) (Sevostyanova *et al.*, 2007). Significant alteration of mRNA expression gradually increased toward the period of host cell lysis (Fig. 23). The numbers of expression-altered genes with q-values of ≤ 0.05 at 25, 50, 75, and 100 min were 0, 15, 269, and 728, respectively, and the numbers of up-regulated and down-regulated genes at 75 min were 144 and 125, and those at 100 min were 332 and 396, respectively.

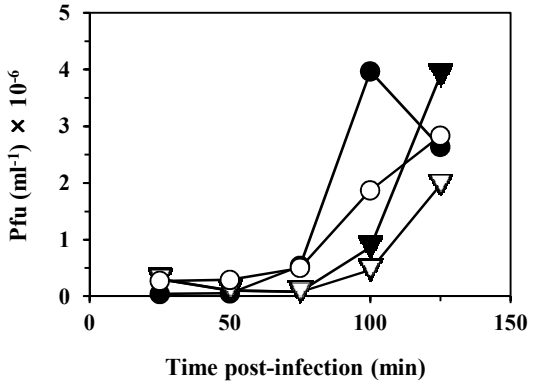


Fig. 22. Progeny phage production in *T. thermophilus* HB8 after infection with phage *Φ*YS40. The wild-type (●,▼) and Δ*crp* (○,▽) strains were infected with *Φ*YS40, and then at each time point post-infection, pfu in the cells (●,○) and in the medium (▼,▽) was determined.

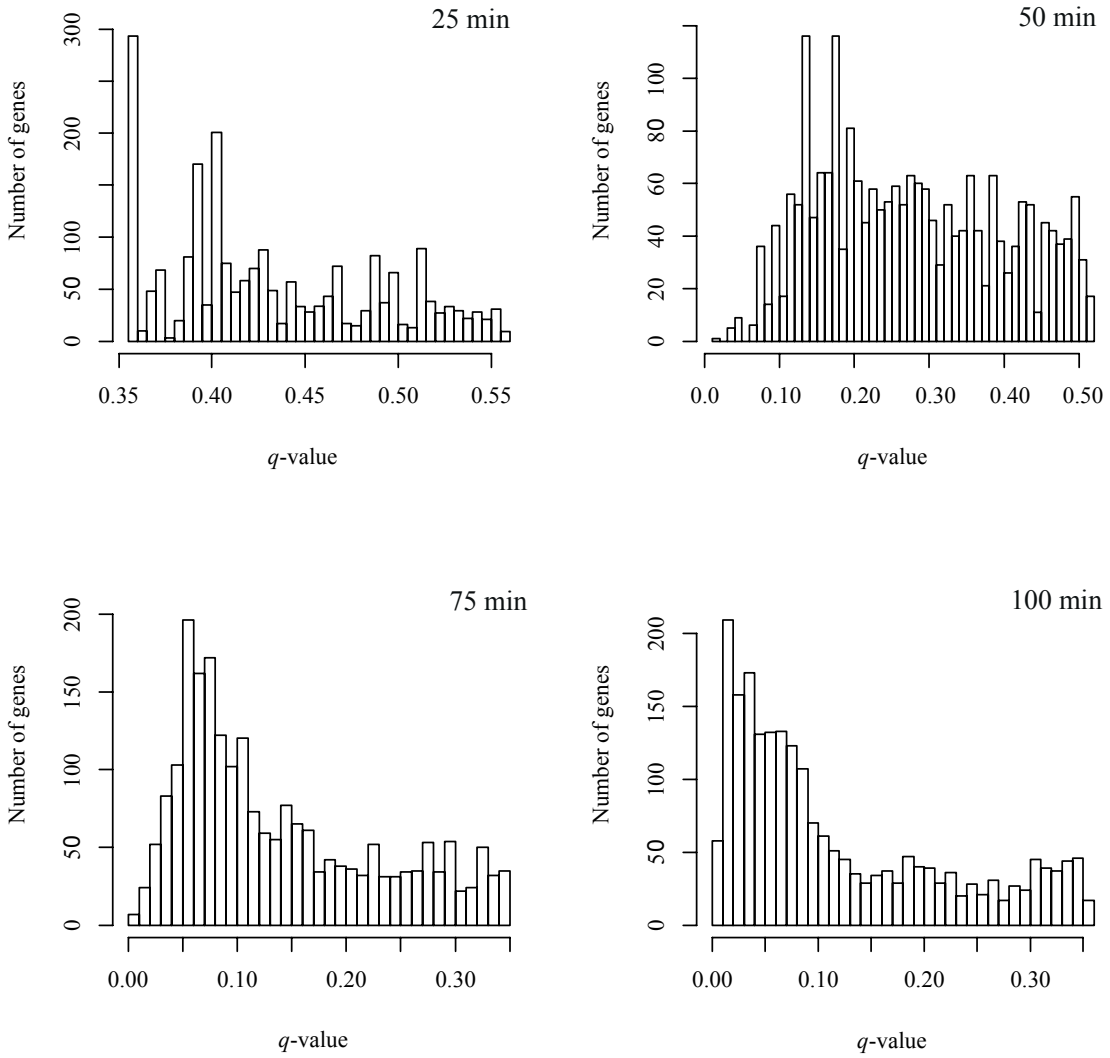


Fig. 23. Alteration of the overall mRNA expression profile in *T. thermophilus* HB8 during infection with phage *Φ*YS40. The numbers of expression-altered genes, with the indicated *q*-values, at 25, 50, 75, and 100 min post-infection, analyzed at the ORF level, are shown.

Expression of *cas* and related genes

Table 12 summarizes the expression levels of representative genes around the CRISPR loci after phage infection relative to those in non-infected cells, as determined on analysis at the ORF level. As a result, the expression of all the genes belonging to the *cas* families, except the *TTHB144–145* (*cas–cas1*) and *TTHB223–224–225* (*cas2–cas1–cas4*) operons and *TTHY7068*

(*cas2*) (positions 136,129–136,401), was found to be significantly up-regulated from around when the phage progeny began to be produced (75 min post-infection). Notably, the *TTHB186–194* operon encoding a set of Ecoli subtype Cas proteins, Cas1–3, and a putative transcriptional regulator were most significantly up-regulated, followed by *TTHB147–152* encoding a set of Mtube subtype and Cas family proteins, both being under the control of a CRP-dependent promoter (Shinkai *et al.*, 2007a). In summary, five core *cas* genes (*cas1*, *cas2*, two *cas3*, and *cas6*), one set each of the Ecoli and Mtube subtypes, RAMP module genes, a putative transcriptional regulator, and a *cas* family gene were significantly expressed ($q \leq 0.05$) after phage infection. Expression of *TTHB068* encoding an argonaute protein, a key catalytic component of the RNA interference pathway (Makarova *et al.*, 2009; Wang *et al.*, 2008), also significantly increased around when host cell lysis began (100 min post-infection). In addition, *TTHB029*, *TTHB175–177*, *TTHB178*, and *TTHB197* encoding a YdjC family protein (Imagawa *et al.*, 2008), putative ion ABC transporter and related proteins, single-stranded DNA-specific exonuclease (Shimada *et al.*, 2010), and formate dehydrogenase, respectively, were significantly increased. Interestingly, many hypothetical genes such as *TTHB028*, *TTHB153*, *TTHB155–159*, *TTHB170*, *TTHB171*, *TTHB184*, *TTHB185*, *TTHB195*, *TTHB198*, *TTHB199*, and *TTHB226–229* were significantly increased after phage infection, especially at 100 min post-infection. Expression of several genes in the vicinity of CRISPRs on the chromosome was not significantly altered after phage infection.

Table 12. Expression of representative genes around the CRISPR loci (see Fig. 19) and the CRP-related genes in the phage-infected *T. thermophilus* HB8 wild-type and Δ*crp* strains at the indicated times post-infection.

Gene name[a]	Annotation for the product	Wild type				Δ*crp*	
		Expression (*q*-value) at time (min)				Expression (*q*-value) at time (min)	
		25	50	75	100	75	100
TTHB028	hypothetical protein	1.33 (0.37)	1.22 (0.14)	1.48 (0.04)	1.77 (0.02)	2.19 (0.02)	2.62 (0.01)
TTHB029	YdjC family protein	1.42 (0.36)	1.67 (0.12)	2.13 (0.01)	2.95 (0.01)	1.60 (0.02)	2.02 (0.00)
TTHB068	argonaute	1.29 (0.39)	1.64 (0.14)	2.06 (0.07)	2.75 (0.02)	1.74 (0.01)	1.72 (0.01)
TTHB144	Cas family protein	1.20 (0.37)	1.33 (0.09)	1.25 (0.14)	1.35 (0.08)	1.76 (0.01)	1.55 (0.01)
TTHB145	Cas1	1.00 (0.55)	0.95 (0.43)	1.00 (0.35)	1.05 (0.29)	1.25 (0.05)	1.29 (0.01)
TTHY7068	Cas2	1.10 (0.46)	0.97 (0.47)	1.18 (0.20)	1.29 (0.08)	1.04 (0.22)	0.93 (0.15)
*TTHB147**	Csm1	1.33 (0.36)	1.98 (0.18)	3.84 (0.01)	4.38 (0.00)	1.09 (0.02)	1.03 (0.09)
*TTHB148**	Csm2	1.42 (0.36)	2.07 (0.16)	3.78 (0.01)	4.60 (0.00)	1.49 (0.02)	1.39 (0.01)
*TTHB149**	Csm3	1.45 (0.36)	1.94 (0.12)	2.66 (0.02)	3.05 (0.01)	1.49 (0.04)	1.55 (0.02)
*TTHB150**	Csm4	1.31 (0.36)	1.89 (0.18)	3.80 (0.02)	4.01 (0.00)	1.77 (0.05)	1.54 (0.05)
*TTHB151**	Csm5	1.30 (0.36)	1.84 (0.18)	3.48 (0.01)	3.93 (0.00)	1.48 (0.05)	1.36 (0.05)
*TTHB152**	Cas family protein	1.37 (0.36)	1.78 (0.13)	2.70 (0.02)	3.09 (0.01)	1.66 (0.03)	1.73 (0.02)
TTHB153	hypothetical protein	1.05 (0.51)	1.16 (0.35)	1.53 (0.08)	1.73 (0.05)	1.57 (0.07)	1.50 (0.05)
TTHB155	hypothetical protein	1.49 (0.40)	1.86 (0.17)	2.12 (0.07)	2.40 (0.05)	2.52 (0.01)	3.04 (0.00)
*TTHB156**	hypothetical protein	2.89 (0.36)	4.19 (0.13)	4.42 (0.06)	5.40 (0.03)	n.d.	n.d.
*TTHB157**	hypothetical protein	2.02 (0.36)	3.16 (0.01)	3.92 (0.00)	4.74 (0.00)	n.d.	n.d.
*TTHB158**	hypothetical protein	1.90 (0.36)	3.04 (0.04)	3.47 (0.03)	4.90 (0.00)	0.98 (0.25)	1.26 (0.11)
*TTHB159**	hypothetical protein	2.12 (0.36)	2.73 (0.10)	2.64 (0.04)	3.86 (0.01)	n.d.	n.d.
TTHB160	Cmr2	1.36 (0.36)	1.91 (0.05)	2.46 (0.02)	2.68 (0.01)	2.28 (0.01)	2.35 (0.00)
TTHB161	Cmr3	1.60 (0.36)	2.10 (0.12)	3.55 (0.03)	3.58 (0.02)	3.25 (0.01)	2.85 (0.00)
TTHB162	Cmr1	1.56 (0.36)	2.04 (0.08)	2.57 (0.03)	3.05 (0.02)	2.40 (0.01)	2.95 (0.01)
TTHB163	Cmr4	1.28 (0.36)	1.74 (0.08)	1.98 (0.08)	2.13 (0.03)	1.82 (0.03)	2.19 (0.01)
TTHB164	Cmr5	1.55 (0.36)	1.80 (0.09)	2.46 (0.05)	2.53 (0.02)	2.26 (0.01)	2.16 (0.00)
TTHB165	Cmr6	1.66 (0.36)	2.05 (0.11)	2.44 (0.04)	2.67 (0.02)	1.90 (0.01)	2.19 (0.00)
TTHB167	hypothetical protein	0.88 (0.51)	0.83 (0.43)	0.68 (0.21)	0.77 (0.26)	0.82 (0.20)	0.85 (0.17)
TTHB168	hypothetical protein	0.94 (0.50)	1.23 (0.23)	1.04 (0.28)	1.09 (0.19)	0.99 (0.24)	1.06 (0.14)
TTHB169	hypothetical protein	1.02 (0.49)	0.98 (0.38)	1.00 (0.34)	0.97 (0.18)	0.96 (0.17)	0.89 (0.03)
TTHB170	hypothetical protein	1.15 (0.40)	1.20 (0.18)	1.27 (0.09)	1.44 (0.03)	0.91 (0.12)	0.92 (0.10)
TTHB171	hypothetical protein	1.25 (0.40)	1.51 (0.16)	1.63 (0.06)	1.86 (0.03)	1.55 (0.02)	1.90 (0.00)
TTHB172	reverse gyrase	1.14 (0.47)	1.02 (0.50)	1.13 (0.26)	1.03 (0.34)	1.74 (0.02)	1.63 (0.01)
TTHB173	response regulator	1.17 (0.42)	1.38 (0.20)	1.74 (0.06)	1.67 (0.06)	1.66 (0.01)	1.70 (0.01)

Continued on next page.

Table 12. Continued.

Gene name[a]	Annotation for the product	Wild type Expression (*q*-value) at time (min)				Δ*crp* Expression (*q*-value) at time (min)	
		25	50	75	100	75	100
TTHB174	sensor histidine kinase-like protein	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
TTHB175	ABC transporter, ATP-binding protein	1.14 (0.47)	1.57 (0.19)	1.43 (0.12)	2.16 (0.04)	1.30 (0.05)	2.00 (0.01)
TTHB176	putative iron ABC transporter, permease protein	1.25 (0.40)	2.07 (0.12)	2.81 (0.05)	4.77 (0.02)	1.34 (0.01)	1.62 (0.03)
TTHB177	iron ABC transporter, periplasmic iron-binding protein	1.27 (0.44)	2.55 (0.16)	3.94 (0.06)	9.52 (0.02)	1.69 (0.02)	2.09 (0.00)
*TTHB178**	single-stranded DNA-specific 3′-5′ exonuclease	1.41 (0.36)	3.02 (0.19)	8.35 (0.01)	9.88 (0.01)	1.00 (0.26)	1.12 (0.06)
TTHB179	hypothetical protein	0.78 (0.36)	0.63 (0.19)	0.67 (0.06)	0.94 (0.26)	1.03 (0.21)	1.15 (0.08)
TTHB180	hypothetical protein	0.86 (0.40)	0.84 (0.15)	0.79 (0.06)	0.90 (0.19)	0.99 (0.25)	1.23 (0.04)
TTHB181	hypothetical protein	0.86 (0.37)	0.80 (0.13)	0.85 (0.13)	1.10 (0.20)	1.01 (0.23)	1.53 (0.03)
TTHB182	hypothetical protein	0.92 (0.40)	0.86 (0.21)	1.02 (0.32)	1.28 (0.06)	1.09 (0.17)	1.88 (0.01)
TTHB183	hypothetical protein	0.93 (0.40)	0.82 (0.17)	0.96 (0.30)	1.17 (0.10)	1.13 (0.15)	1.51 (0.02)
TTHB184	hypothetical protein	1.60 (0.36)	1.49 (0.19)	1.44 (0.11)	1.97 (0.04)	1.60 (0.03)	1.81 (0.00)
TTHB185	hypothetical protein	1.60 (0.36)	1.60 (0.12)	1.78 (0.05)	2.40 (0.02)	2.01 (0.02)	2.44 (0.01)
*TTHB186**	predicted transcriptional regulator	2.01 (0.36)	4.04 (0.10)	8.39 (0.03)	9.51 (0.02)	1.52 (0.02)	2.23 (0.00)
*TTHB187**	Cas3	2.16 (0.37)	4.87 (0.13)	9.36 (0.04)	10.7 (0.03)	1.66 (0.05)	2.27 (0.02)
*TTHB188**	Cse1	1.63 (0.36)	2.84 (0.12)	5.67 (0.03)	7.46 (0.02)	2.12 (0.03)	2.82 (0.01)
*TTHB189**	Cse2	1.41 (0.36)	2.46 (0.13)	4.53 (0.03)	5.78 (0.01)	1.74 (0.01)	2.15 (0.01)
*TTHB190**	Cse4	1.57 (0.36)	3.00 (0.14)	6.04 (0.02)	7.53 (0.01)	2.56 (0.01)	2.85 (0.01)
*TTHB191**	Cas5e	1.65 (0.36)	3.01 (0.12)	5.63 (0.01)	7.19 (0.01)	2.09 (0.02)	2.63 (0.01)
*TTHB192**	Cse3	1.55 (0.36)	2.99 (0.14)	6.47 (0.02)	8.02 (0.01)	2.20 (0.02)	2.91 (0.01)
*TTHB193**	Cas1	1.62 (0.36)	2.69 (0.13)	6.06 (0.03)	6.95 (0.02)	2.46 (0.02)	2.71 (0.01)
*TTHB194**	Cas2	1.46 (0.36)	2.38 (0.15)	4.37 (0.01)	5.57 (0.01)	2.48 (0.02)	2.80 (0.01)
TTHB195	hypothetical protein	1.17 (0.39)	1.23 (0.17)	1.29 (0.07)	1.31 (0.05)	1.66 (0.01)	2.01 (0.00)
TTHB196	putative protein required for formate dehydrogenase activity	1.05 (0.51)	0.88 (0.41)	1.09 (0.28)	1.54 (0.07)	1.09 (0.16)	1.59 (0.01)
TTHB197	formate dehydrogenase	1.29 (0.40)	1.82 (0.12)	2.78 (0.04)	3.71 (0.01)	2.64 (0.01)	2.96 (0.00)
TTHB198	hypothetical protein	1.52 (0.37)	2.23 (0.12)	3.23 (0.03)	4.43 (0.01)	1.81 (0.01)	1.76 (0.01)
TTHB199	hypothetical protein	1.54 (0.36)	2.06 (0.12)	2.94 (0.04)	3.93 (0.01)	2.21 (0.02)	2.44 (0.01)
TTHB202	hypothetical protein	1.08 (0.49)	0.90 (0.45)	0.67 (0.10)	0.90 (0.25)	0.67 (0.05)	0.86 (0.12)
TTHB222	hypothetical protein	0.89 (0.50)	0.96 (0.49)	0.90 (0.29)	0.86 (0.27)	0.72 (0.08)	0.66 (0.02)
TTHB223	Cas2	1.38 (0.40)	1.26 (0.31)	1.27 (0.18)	1.42 (0.12)	0.78 (0.10)	1.47 (0.05)
TTHB224	Cas1	1.10 (0.48)	1.05 (0.48)	0.74 (0.15)	1.23 (0.18)	1.17 (0.20)	1.56 (0.09)
TTHB225	Cas4	1.22 (0.45)	1.20 (0.39)	1.03 (0.33)	1.36 (0.19)	1.38 (0.13)	1.88 (0.05)

Continued on next page.

Table 12. Continued.

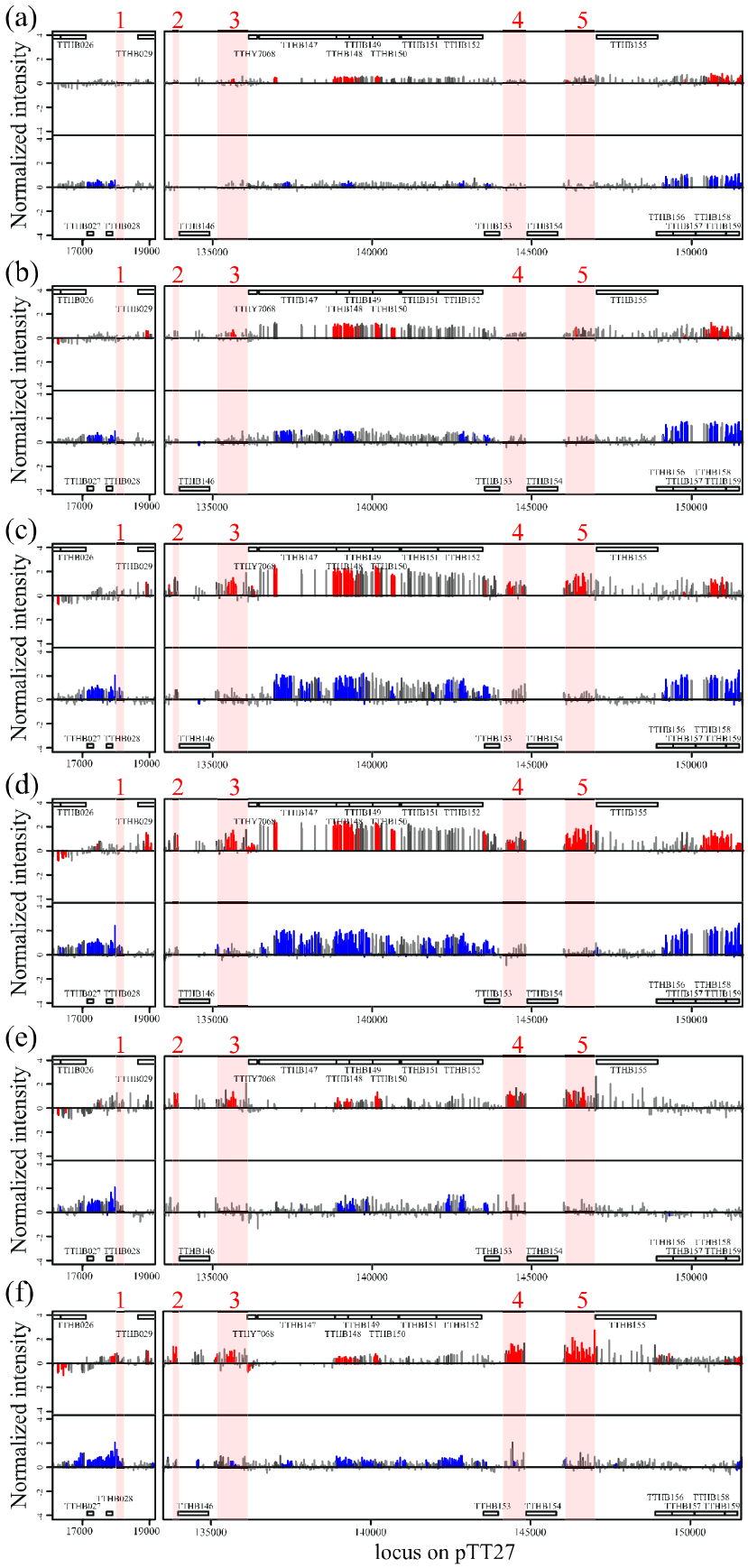
Gene name[a]	Annotation for the product	Wild type Expression (*q*-value) at time (min)				Δ*crp* Expression (*q*-value) at time (min)	
		25	50	75	100	75	100
TTHB226	hypothetical protein	1.24 (0.41)	1.67 (0.10)	2.61 (0.04)	3.19 (0.01)	1.81 (0.03)	2.23 (0.01)
TTHB227	hypothetical protein	1.05 (0.40)	1.40 (0.10)	1.65 (0.03)	1.98 (0.02)	1.41 (0.04)	1.43 (0.03)
TTHB228	hypothetical protein	1.11 (0.36)	1.40 (0.13)	1.98 (0.01)	2.80 (0.01)	2.01 (0.01)	2.28 (0.00)
TTHB229	hypothetical protein	1.28 (0.36)	1.70 (0.13)	2.40 (0.04)	3.46 (0.01)	2.18 (0.01)	2.57 (0.00)
TTHB230	Cas3	1.17 (0.36)	1.41 (0.08)	1.76 (0.05)	2.47 (0.01)	2.06 (0.01)	2.15 (0.01)
TTHB231	Cas6	1.14 (0.40)	1.29 (0.17)	1.80 (0.05)	2.14 (0.03)	1.59 (0.01)	1.80 (0.01)
TTHB233	hypothetical protein	1.15 (0.39)	1.14 (0.29)	1.06 (0.24)	0.99 (0.33)	1.04 (0.22)	0.80 (0.05)
*TTHA0176**	GCN5-related acetyltransferase	0.87 (0.37)	0.85 (0.18)	1.07 (0.18)	1.08 (0.15)	0.71 (0.01)	0.72 (0.01)
*TTHA0771**	transcriptional activator SARP family	1.81 (0.36)	3.85 (0.08)	12.4 (0.02)	22.0 (0.01)	4.21 (0.01)	6.06 (0.01)
TTHA0798	GGDEF domain protein	1.28 (0.36)	1.81 (0.15)	3.33 (0.02)	4.36 (0.01)	1.63 (0.02)	2.46 (0.00)
TTHA0909	ornithine aminotransferase	1.08 (0.47)	1.02 (0.49)	0.96 (0.30)	0.99 (0.35)	1.00 (0.25)	0.98 (0.19)
TTHA0910	exoribonuclease	1.11 (0.43)	1.04 (0.43)	1.06 (0.27)	0.96 (0.30)	1.07 (0.17)	0.91 (0.10)
TTHA0924	hypothetical protein	1.07 (0.44)	1.04 (0.40)	1.08 (0.07)	1.01 (0.34)	0.83 (0.05)	0.87 (0.06)
TTHA0925	poly(A) polymerase family protein	0.79 (0.40)	0.99 (0.50)	1.25 (0.12)	1.32 (0.09)	0.84 (0.04)	0.66 (0.01)
TTHA1437	CRP	1.08 (0.47)	1.14 (0.31)	0.95 (0.30)	0.93 (0.24)	n.d.	n.d.

Normalized intensities in the post-infected cells relative to those in the non-infected cells and the *q*-values, determined at the ORF level, are shown.
n.d., the detection call was absent (see Materials and Methods).
[a]Genes with asterisks are under the control of CRP-dependent promoters. Expression of a probable transposase gene (see Fig. 19) could not be determined because oligonucleotide probes for it correspond to a couple of the same or similar genes. Expression of *TTHB166* gene cannot be determined because adequate probes cannot be designed.

Expression of CRISPRs

In order to investigate the altered expression of CRISPRs after phage infection, probe-level analysis of the DNA microarray data was performed. After phage infection, the normalized intensities of most probes corresponding to one strand each of all CRISPRs except CRISPR-8 and CRISPR-10, which were not detected because corresponding oligonucleotide probes have not been designed for DNA microarray analysis, gradually increased with *p*-values of $\leq 10^{-4}$; thus, the transcripts derived from the CRISPRs were mostly generated from one strand on phage

infection (Figs. 19 and 24). The expression levels differed among the loci, roughly in the order of CRISPR-5, -6, -7, -9, and -11 > CRISPR-3, -4, and -12 > CRISPR-1 and -2 (Fig. 24), which is not the same as that of the associated *cas* genes (Table 12 and Fig. 19). Highly conserved regions were found in the upstream sequences of CRISPR-1, -2, -4, and -11, and between those of CRISPR-6 and -7, respectively (Fig. 25). These observations support the transcriptional direction of the CRISPRs determined on DNA microarray analysis because upstream sequences of the CRISPRs possibly contained promoters for the following CRISPRs (Brouns *et al.*, 2008). The upstream sequences of CRISPR-9 and -10 are also conserved (Fig. 25). The repeat sequences of the CRISPRs were classified into three types: those of CRISPR-1, -2, -3, -4, -5, -8, -11, and -12 (type I); those of CRISPR-6 and -7 (type II); and those of CRISPR-9 and -10 (type III) (Table 13). CRISPR-3, -5, and -12 had possibly been generated through insertion of the same or a similar transposase gene, *TTHB146*, *TTHB154*, and *TTHA0921*, respectively (Fig. 19). The increased expression level did not depend on the type of repeat sequence.



Continued on next page.

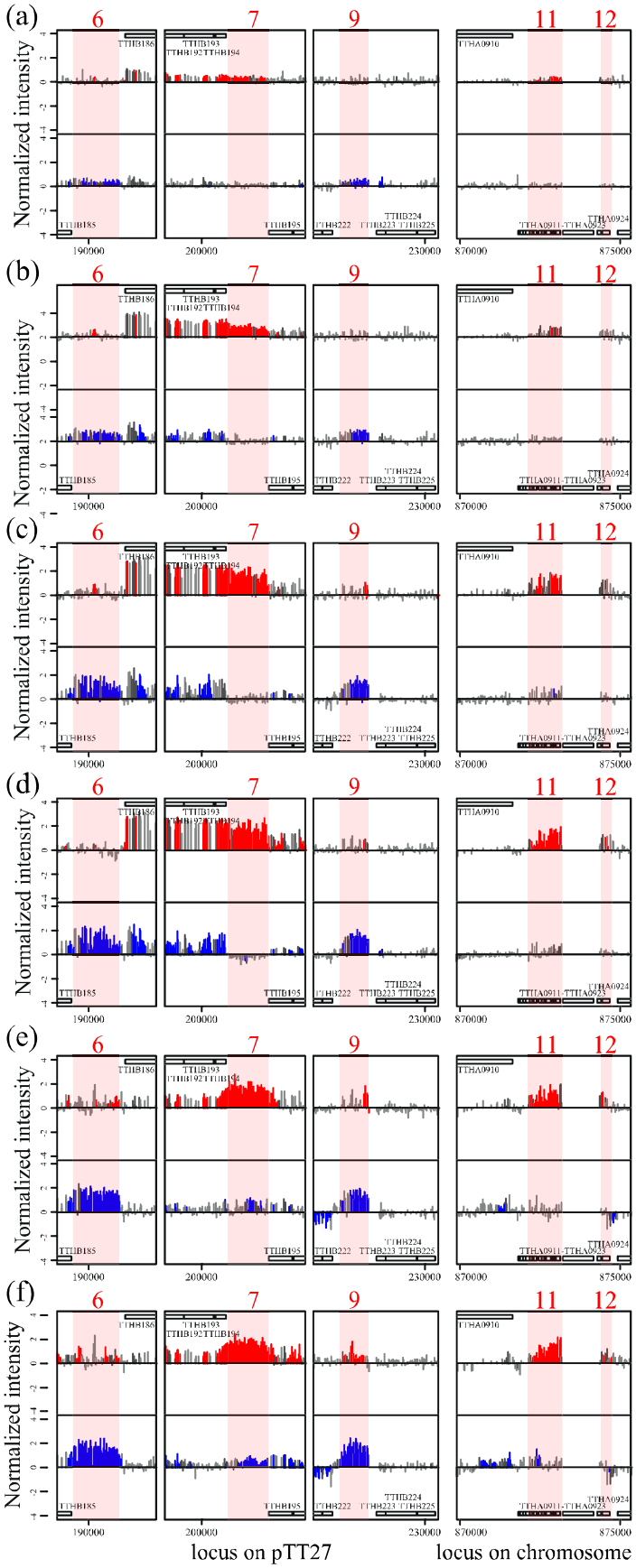


Fig. 24. Altered expression around the CRISPR loci after phage infection. The expression in the phage-infected *T. thermophilus* HB8 wild-type strain at 25 **(a)**, 50 **(b)**, 75 **(c)**, and 100 min **(d)** and the Δ*crp* strain at 75 **(e)** and 100 min **(f)** post-infection, relative to those in the non-infected strains, was analyzed at the probe level, and is presented as the $\log_2$-transformed normalized intensity of each probe. The upper and lower graphs for each time point are for the (+) and (−) strands, respectively. Numerals indicated by red letters on the graphs represent the CRISPR code numbers. CRISPR regions are indicated in red on the genome map. The *p*-value range is indicated by a color; that is, $p \leq 10^{-4}$ for the (+) strand (red), $p \leq 10^{-4}$ for the (−) strand (blue), $10^{-4} < p \leq 10^{-3}$ (black), and $p < 10^{-3}$ (gray). The expression of CRISPR-8 and -10 was not detected, because the corresponding oligonucleotide probes have not been designed for DNA microarray analyses.

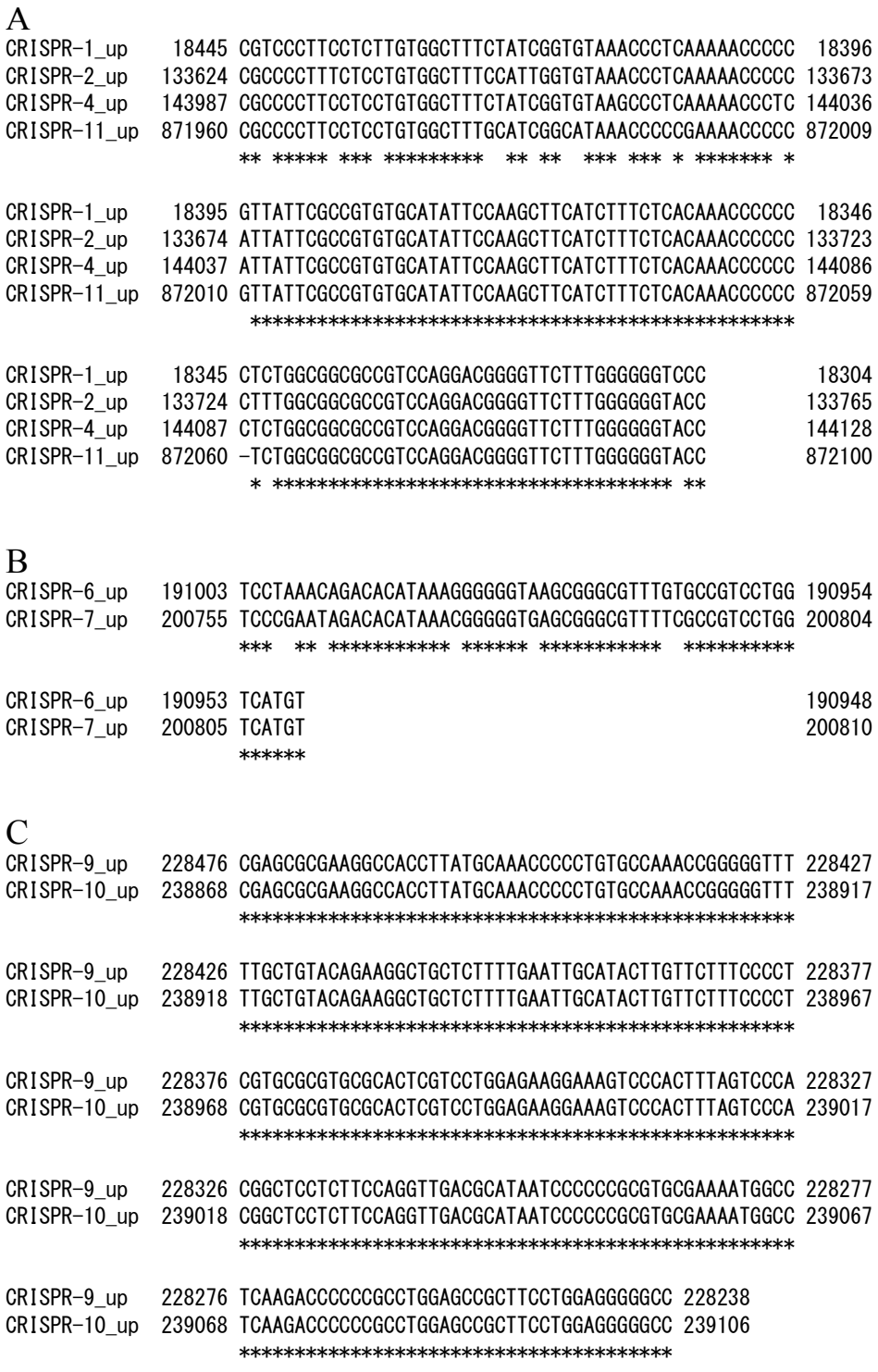


Fig. 25. Nucleotide sequence alignment of upstream regions of CRISPR-1, -2, -4, and -11 **(A)** CRISPR-6, and -7 **(B)** and CRISPR-9, and -10 **(C)**. Conserved sequences are indicated by asterisks. The genome positions of the sequences on pTT27 (CRISPR-1, -2, -4, -6, -7, -9, and -10) and the chromosome (CRISPR-11) are indicated.

Table 13. Composition of the CRISPR loci on pTT27 and the chromosome of *T. thermophilus* HB8.

Locus[a]	CRISPR name	Representative repeat sequence (5′ to 3′)[b]	Type of repeat sequence	No. of repeats	Length of spacer sequence (avr) (bp)
on pTT27					
18,303–18,044	CRISPR-1			4	38–40 (38.7)
133,766–133,954	CRISPR-2			3	40–41 (40.5)
135,156–136,099	CRISPR-3	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	13	35–42 (39.6)
144,129–144,842	CRISPR-4			10	37–42 (39.3)
146,042–146,983	CRISPR-5			13	36–44 (39.5)
190,947–189,507	CRISPR-6	GTAGTCCCCACGCACGTGGGGATGGACCG	II	24	31–33 (32.4)
200,811–202,078	CRISPR-7	GTAGTCCCCACGCGTGTGGGGATGGACCG		21	32–35 (33.0)
210,807–210,842	CRISPR-8	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGATAC	I	1	-
228,237–227,324	CRISPR-9	GTTGCAAACCCCGTCAGCCTCGTAGAGGATTGAAAC	III	13	36–41 (37.2)
239,108–239,214	CRISPR-10	GTTGCAAACCTCGTTAGCCTCGTAGAGGATTGAAAC		2	35
On the chromosome					
872,101–873,199	CRISPR-11	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	15	35–43 (39.9)
874,397–874,734	CRISPR-12			5	37–41 (39.5)

[a]Each one begins with the first base of the first repeat, and ends with the last base of the last repeat.
[b]Palindromic sequences are underlined.

Expression of CRP-related genes

T. thermophilus CRP positively regulates 22 genes in a cAMP-dependent manner (Shinkai *et al.*, 2007a). In addition to two CRP-regulated *cas* operons (Fig. 19), the expression of the remaining CRP-regulated genes except *TTHA0176*, remarkably increased after phage infection, being especially increased at 100 min post-infection (Table 12). Increased expression of the *crp* gene was not observed; furthermore, expression of the CRP-regulated genes was less up-regulated in a phage-infected Δ*crp* strain than in the wild type (Table 12; see "Effects of *crp* gene disruption on responses to phage infection" subsection). These results suggest that expression of these CRP-regulated genes increases due to an increase in the intracellular cAMP

level, and that cAMP is a signaling molecule that transmits information on phage infection to CRP. An additional regulatory factor might be involved in the expression of *TTHA0176*. Interestingly, the expression of *TTHA0798* encoding a GGDEF domain protein, which has a domain homologous to an adenylyl cyclase catalytic one, also increased after phage infection (Table 12). This protein might be involved in cAMP synthesis. The increased cAMP level might be undetectable because it could not be observed upon enzyme immunoassaying with anti-cAMP antibodies, performed as described previously (Shinkai *et al.*, 2007a).

Several other features observed in the genome-wide expression profile

In order to characterize the genome-wide altered expression profile, I categorized all of the expression-altered genes ($q \leq 0.05$) at 75 and 100 min post-infection based on the Clusters of Orthologous Groups of proteins (COGs) (http://www.ncbi.nlm.nih.gov/COG) (Tatusov *et al.*, 2003), as summarized in Table 14 and Fig. 26 (Agari *et al.*, 2010b). In category T, the number of up-regulated genes was more than that of down-regulated ones (Table 14). The relative expression levels of the up-regulated genes at 75 and 100 min were 1.38–12.4 [average value (avr), 3.84] and 1.28–22.0 (avr, 3.29), respectively. Interestingly, 6 putative two-component response regulators out of the 12 found in this strain were up-regulated 1.29–2.76 (avr, 1.89) times at 100 min (Agari *et al.*, 2010b). On the other hand, many down-regulated genes were found in categories J, E, F, H, and I, especially at 100 min (Table 14). Their relative

expression levels were 0.48–0.88 (avr, 0.61), 0.33–0.81 (avr, 0.64), 0.36–0.91 (avr, 0.67), 0.29–0.86 (avr, 0.62), and 0.41–0.84 (avr, 0.67), respectively. Notably, the expression of the genes in category L and that of uncategorized genes, including several functionally uncharacterized *cas* genes, were relatively greatly altered (Fig. 26), with more than 70% of them being up-regulated 1.71–8.35 (avr, 3.50) and 1.17–6.47 (avr, 2.52) times, respectively, at 75 min [Table 14, and see also supplementary Table S1 of ref. (Agari *et al.*, 2010b)]. As reported previously (Sevostyanova *et al.*, 2007), the expression of *rpoC* (*TTHA1812*) was decreased at 100 min, the relative expression level being 0.57 (q = 0.04). As for *Φ*YS40 genes, expression of the genes for DNA replication, recombination, and nucleotide metabolism increases from the early stage of post-infection (Sevostyanova *et al.*, 2007). Taking all the facts together, I speculate that upon phage infection, a signal is transmitted to the host cells, and the overall efficiencies of transcription, translation, and metabolism in the cells decrease, while replication of phage DNA and uncharacterized host response systems including CRISPR systems (see above) are activated. Unlike in the previous study (Sevostyanova *et al.*, 2007), significantly altered expression of *infB* (*TTHA0699*), *infC* (*TTHA0551*), *sigA* (*TTHA0532*), *dnaK* (*TTHA1491*), and the alcohol dehydrogenase gene (*TTHA0466*) was not observed.

Table 14. Number of expression-altered genes ($q \leq 0.05$, on ORF-level analysis) in each category of COGs observed in the phage-infected *T. thermophilus* HB8 wild-type and Δ*crp* strains at 75 and 100 min post-infection.

COG code	Description	In genome[a]	Wild Type				Δ*crp*			
			75 min		100 min		75 min		100 min	
			Up (%)	Down (%)	Up (%)	Down (%)	Up (%)	Down (%)	Up (%)	Down (%)
Information storage and processing										
J	Translation	153	2 (15.4)	11 (84.6)	4 (14.3)	24 (85.7)	10 (8.3)	111 (91.7)	8 (6.6)	113 (93.4)
K	Transcription	104	9 (56.3)	7 (43.7)	14 (41.1)	20 (58.9)	29 (46.8)	33 (53.2)	39 (50.6)	38 (49.4)
L	Replication, recombination and repair	115	14 (77.8)	4 (22.2)	21 (56.8)	16 (43.2)	33 (49.2)	34 (50.8)	41 (54.7)	34 (45.3)
B	Chromatin structure and dynamics	2	0 (0)	0 (0)	0 (0)	1 (100)	0 (0)	0 (0)	0 (0)	0 (0)
Cellular processes										
D	Cell cycle control, mitosis and meiosis	28	1 (50.0)	1 (50.0)	2 (22.2)	7 (77.8)	8 (53.3)	7 (46.7)	7 (43.8)	9 (56.2)
V	Defense mechanisms	20	1 (33.3)	2 (66.7)	2 (33.3)	4 (66.7)	3 (42.9)	4 (57.1)	6 (54.5)	5 (45.5)
T	Signal transduction mechanisms	65	6 (85.7)	1 (14.3)	16 (84.2)	3 (15.8)	31 (86.1)	5 (13.9)	38 (86.4)	6 (13.6)
M	Cell wall/membrane biogenesis	78	5 (50.0)	5 (50.0)	10 (31.3)	22 (68.7)	14 (31.1)	31 (68.9)	23 (37.7)	38 (62.3)
N	Cell motility	12	1 (50.0)	1 (50.0)	2 (50.0)	2 (50.0)	4 (44.4)	5 (55.6)	3 (37.5)	5 (62.5)
Z	Cytoskeleton	1	1 (100)	0 (0)	1 (100)	0 (0)	1 (100)	0 (0)	1 (100)	0 (0)
U	Intracellular trafficking and secretion	30	1 (33.3)	2 (66.7)	2 (20.0)	8 (80.0)	5 (27.8)	13 (72.2)	6 (30.0)	14 (70.0)
O	Posttranslational modification, protein turnover, and chaperones	84	2 (28.6)	5 (71.4)	13 (48.1)	14 (51.9)	8 (20.0)	32 (80.0)	15 (30.0)	35 (70.0)
Metabolism										
C	Energy production and conversion	150	11 (47.8)	12 (52.2)	19 (45.2)	23 (54.8)	42 (61.8)	26 (38.2)	59 (60.8)	38 (39.2)
G	Carbohydrate transport and metabolism	125	11 (68.8)	5 (31.2)	26 (57.8)	19 (42.2)	34 (47.9)	37 (52.1)	43 (51.2)	41 (48.8)
E	Amino acid transport and metabolism	209	12 (70.6)	5 (29.4)	25 (39.0)	39 (61.0)	48 (38.4)	77 (61.6)	54 (36.7)	93 (63.3)
F	Nucleotide transport and metabolism	64	0 (0)	5 (100)	0 (0)	26 (100)	7 (16.3)	36 (83.7)	11 (24.4)	34 (75.6)
H	Coenzyme transport and metabolism	110	4 (25.0)	12 (75.0)	13 (33.3)	26 (66.7)	22 (33.3)	44 (66.7)	36 (40.9)	52 (59.1)
I	Lipid transport and metabolism	89	4 (33.3)	8 (66.7)	13 (37.1)	22 (62.9)	24 (45.3)	29 (54.7)	27 (45.0)	33 (55.0)
P	Inorganic ion transport and metabolism	91	3 (50.0)	3 (50.0)	14 (63.6)	8 (26.4)	18 (32.7)	37 (67.3)	20 (33.3)	40 (66.7)
Q	Secondary metabolites biosynthesis, transport and catabolism	58	1 (16.7)	5 (83.3)	12 (52.1)	11 (47.9)	21 (65.6)	11 (34.4)	27 (71.1)	11 (28.9)
Poorly characterized										
R	General function prediction only	304	19 (55.9)	15 (44.1)	47 (45.6)	56 (54.4)	66 (39.1)	103 (60.9)	89 (41.8)	124 (58.2)
S	Function unknown	166	15 (75.0)	5 (25.0)	26 (45.6)	31 (54.4)	42 (50.0)	42 (50.0)	47 (46.5)	54 (53.5)
-	Not in COGs	434	46 (70.8)	19 (29.2)	93 (60.4)	61 (39.6)	126 (58.9)	88 (41.1)	187 (63.6)	113 (36.4)
	Total	2492	169 (56.0)	133 (44.0)	375 (45.8)	443 (54.2)	596 (42.5)	805 (57.5)	787 (45.8)	930 (54.2)

Up and Down indicate up-regulated and down-regulated genes, respectively. The percentage of the altered genes in each category is shown in parentheses.

[a]Number of genes in the genome of *T. thermophilus* HB8.

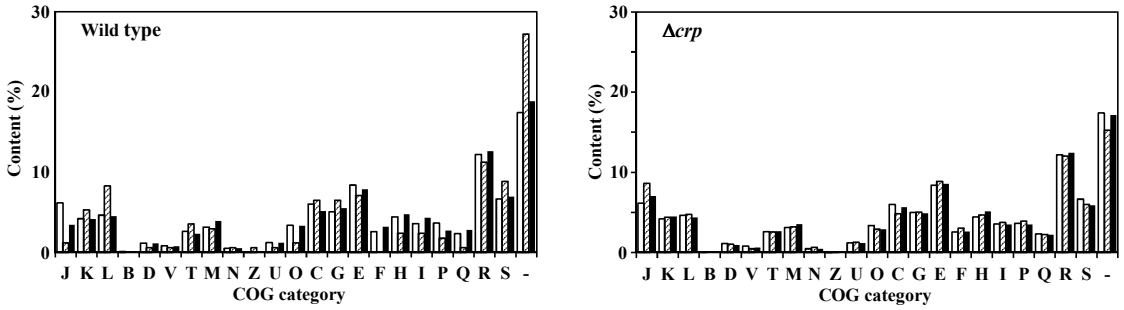


Fig. 26. Contents of expression-altered genes in the phage *Φ*YS40-infected *T. thermophilus* HB8 wild-type and Δ*crp* strains. The percentages of each COG-based categorized gene among the total number of expression-altered genes ($q \leq 0.05$) at 75 (shaded bars) and 100 min (dark bars) post-infection are shown. The percentages of the genes in the genome are shown by open bars. The definition for each COG code is given in Table 14.

Effects of *crp* gene disruption on responses to phage infection

The expression of CRP-regulated genes was most remarkably up-regulated by phage infection (Table 12); thus CRP may be one of the key molecules for cells to respond to phage infection. In order to determine the effects of this molecule on host responses after phage infection, I infected the Δ*crp* strain with *Φ*YS40. The eclipse period of the strain was almost the same as that of the wild type, suggesting that CRP-regulated genes are not involved in phage production (Fig. 20).

The expression of the *TTHB159–156* operon and *TTHB178*, which are directly controlled by CRP, was not up-regulated in the phage-infected Δ*crp* strain (Table 12). In the case of the two *cas* operons and *TTHA0771*, which are under the control of CRP-dependent promoters (Fig. 19 and Table 12), increased expression was still observed in the Δ*crp* strain after phage infection, although the level was only 25–50% of that in the phage-infected wild-type strain (Table 12), suggesting that their expression is controlled by both CRP and some additional regulatory

factor(s). Expression of *TTHB176* and *TTHB177* was less up-regulated even though they are not under the control of a CRP-dependent promoter, suggesting that their expression is indirectly controlled by CRP. In the case of *cas* genes that are not under the control of a CRP-dependent promoter, [i.e., the *TTHB160–165* operon encoding a set of RAMP modules, *TTHB230* (*cas3*), and *TTHB231* (*cas6*)], increased expression was observed in the phage-infected Δ*crp* strain, with the level being similar to that in the phage-infected wild type (Table 12). For up-regulation of these genes, unknown regulatory factor(s) may be activated when a phage infects cells. Additionally, expression of the *TTHB144–145* (*cas–cas1*) operon was significantly up-regulated in the Δ*crp* strain after phage infection, unlike in the wild type (Table 12). Expression of all remaining *cas* genes was not altered in the phage-infected Δ*crp* strain, as observed in the wild type (Table 12). Regarding the expression of the *cas* genes, basically the same results were obtained on probe-level analysis of the DNA microarray data (Agari *et al.*, 2010b). Thus, the transcriptional regulatory mechanism may differ depending on the *cas* gene. The up-regulation of the transcription of CRISPRs in the phage-infected Δ*crp* strain was observed to be similar to that in the wild type (Fig. 24), indicating that the regulation was not mediated by CRP, but by unknown regulatory factor(s) induced by phage infection.

As in the case of the wild-type strain, significant alteration of overall mRNA expression was observed in the Δ*crp* strain (Agari *et al.*, 2010b). Although the numbers of up-regulated and down-regulated genes ($q \leq 0.05$) differed from those in the wild type, the expression profiles of

the genes involved in signal transduction, transcription, translation, and metabolism were basically consistent with those in the wild type. However, the ratio of the genes belonging to category L and the uncategorized genes among the expression-altered genes at 75 min post-infection was reduced in the Δ*crp* strain, which was the most striking feature of the profile (Fig. 26). This finding suggests that unknown host response systems as well as several CRISPR systems are under the control of CRP.

2-5. Discussion

I found that most CRISPR systems of *T. thermophilus* HB8 were significantly up-regulated by phage infection. Interestingly, expression of the systems was not simply regulated by some common regulatory factor(s); that is, several *cas* genes were regulated by CRP, which may be the receiver of a signaling molecule, cAMP, while several other *cas* genes and CRISPRs were regulated by unidentified factor(s). Notably, CRISPR and the genes in the vicinity of the CRISPRs including *cas* genes were not always simultaneously up-regulated after phage infection, as in the cases of the CRISPR-2, -3, -11, and -12 loci (Fig. 19). It might be that the protein machineries of CRISPR systems can interact in *trans* with RNA transcribed from distant CRISPRs. Several core *cas* genes were not up-regulated after phage infection, but were expressed in the non-infected strains because the detection calls for them were all 'present'. Interestingly, up-regulation of only one each of the *cas1* and *cas2* genes was observed although the strain has two more of each gene. It would be more interesting if the expression profile of *cas* genes could be altered depending on the species of invading foreign replicons. Since the bacterial strain used in this study is sensitive to phage *Φ*YS40 and the strain does not have CRISPR spacers identical with its DNA, the phage-induced expression profiles I observed might reflect the adaptation phase rather than the interference phase of CRISPR immunity. Alternatively, the strain might respond similarly to any foreign replicon, and it might be ready to adapt and interfere with its infection simultaneously even if it does not have CRISPR spacers

identical with the sequences of the invading foreign replicon because Cas1, which has been thought to be involved in the adaptation phase (Wiedenheft *et al.*, 2009), and Cas6 (Carte *et al.*, 2010) and Ecoli subtype Cas proteins plus Cas3 (Brouns *et al.*, 2008), which have been thought to be involved in the interference phase, were significantly up-regulated after phage infection. I found that about 40% of the expression-altered genes of wild-type *T. thermophilus* HB8 after phage infection ($q \leq 0.05$) are annotated as being functionally unknown, suggesting the involvement of unknown host response system(s) in addition to CRISPR systems. Of the functionally unknown genes, significantly up-regulated genes after phage infection, such as *TTHB028*, *TTHB153*, *TTHB155–159*, *TTHB184*, *TTHB185*, *TTHB195*, *TTHB198*, *TTHB199*, and *TTHB226–229*, which are localized in close proximity in CRISPRs (Fig. 19), might play key roles in the CRISPR immunity.

Thus, the results obtained in this study provide a basis for additional biochemical and genetic characterization of host response systems including CRISPR systems, which will facilitate understanding of the regulatory mechanism induced by phage infection. Since proteins from *T. thermophilus* HB8 are good candidates for X-ray crystal structure analysis (Yokoyama *et al.*, 2000a), this results will also facilitate understanding of the mechanism at the atomic level.

References

Agari, Y., Kashihara, A., Yokoyama, S., Kuramitsu, S. & Shinkai, A. (2008a). Global gene expression mediated by *Thermus thermophilus* SdrP, a CRP/FNR family transcriptional regulator. *Mol Microbiol* **70**, 60-75.

Agari, Y., Yokoyama, S., Kuramitsu, S. & Shinkai, A. (2008b). X-ray crystal structure of a CRISPR-associated protein, Cse2, from *Thermus thermophilus* HB8. *Proteins* **73**, 1063-1067.

Agari, Y., Kuramitsu, S. & Shinkai, A. (2010a). Identification of novel genes regulated by the oxidative stress-responsive transcriptional activator SdrP in *Thermus thermophilus* HB8. *FEMS Microbiol Lett* **313**, 127-134.

Agari, Y., Sakamoto, K., Tamakoshi, M., Oshima, T., Kuramitsu, S. & Shinkai, A. (2010b). Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. *J Mol Biol* **395**, 270-281.

Alfonso, M., Perewoska, I. & Kirilovsky, D. (2001). Redox control of *ntcA* gene expression in *Synechocystis* sp. PCC 6803. Nitrogen availability and electron transport regulate the levels of the NtcA protein. *Plant Physiol* **125**, 969-981.

Almeida, M. S., Herrmann, T., Peti, W., Wilson, I. A. & Wüthrich, K. (2005). NMR structure of the conserved hypothetical protein TM0487 from *Thermotoga maritima*: implications for 216 homologous DUF59 proteins. *Protein Sci* **14**, 2880-2886.

Anjum, M. F., Green, J. & Guest, J. R. (2000). YeiL, the third member of the CRP-FNR

family in *Escherichia coli*. *Microbiol* **146 Pt 12**, 3157-3170.

Bai, G., McCue, L. A. & McDonough, K. A. (2005). Characterization of *Mycobacterium tuberculosis* Rv3676 (CRPMt), a cyclic AMP receptor protein-like DNA binding protein. *J Bacteriol* **187**, 7795-7804.

Ballesteros, M., Fredriksson, Å., Henriksson, J. & Nyström, T. (2001). Bacterial senescence: protein oxidation in non-proliferating cells is dictated by the accuracy of the ribosomes. *EMBO J* **20**, 5280-5289.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712.

Beloglazova, N., Brown, G., Zimmerman, M. D. & other authors (2008). A novel family of sequence-specific endoribonucleases associated with the Clustered Regularly Interspaced Short Palindromic Repeats. *J Biol Chem* **283**, 20361-20371.

Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiol* **151**, 2551-2561.

Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-193.

Botsford, J. L. & Harman, J. G. (1992). Cyclic AMP in prokaryotes. *Microbiol Rev* **56**, 100-122.

Brünger, A. T., Adams, P. D., Clore, G. M. & other authors (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-921.

Bridges, B. A. (1993). Spontaneous mutation in stationary-phase *Escherichia coli* WP2 carrying various DNA repair alleles. *Mutat Res* **302**, 173-176.

Brouns, S. J., Jore, M. M., Lundgren, M. & other authors (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964.

Browning, D. F. & Busby, S. J. (2004). The regulation of bacterial transcription initiation. *Nature Rev Microbiol* **2**, 57-65.

Busby, S. & Ebright, R. H. (1999). Transcription activation by catabolite activator protein (CAP). *J Mol Biol* **293**, 199-213.

Cameron, A. D. & Redfield, R. J. (2006). Non-canonical CRP sites control competence regulons in *Escherichia coli* and many other gamma-proteobacteria. *Nucleic Acids Res* **34**, 6001-6014.

Carte, J., Pfister, N. T., Compton, M. M., Terns, R. M. & Terns, M. P. (2010). Binding and cleavage of CRISPR RNA by Cas6. *RNA* **16**, 2181-2188.

Cawley, S., Bekiranov, S., Ng, H. H. & other authors (2004). Unbiased mapping of

transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499-509.

Cheng Vollmer, A. & Van Dyk, T. K. (2004). Stress responsive bacteria: biosensors as environmental monitors. *Adv Microb Physiol* **49**, 131-174.

Chibani-Chennoufi, S., Bruttin, A., Dillmann, M. L. & Brussow, H. (2004). Phage-host interaction: an ecological perspective. *J Bacteriol* **186**, 3677-3686.

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763.

Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190.

Dai, M., Wang, P., Boyd, A. D. & other authors (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res* **33**, e175.

Derouaux, A., Dehareng, D., Lecocq, E. & other authors (2004). Crp of *Streptomyces coelicolor* is the third transcription factor of the large CRP-FNR superfamily able to bind cAMP. *Biochem Biophys Res Commun* **325**, 983-990.

Ebihara, A., Yao, M., Masui, R., Tanaka, I., Yokoyama, S. & Kuramitsu, S. (2006). Crystal structure of hypothetical protein TTHB192 from *Thermus thermophilus* HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein Sci* **15**, 1494-1499.

Ebright, R. H., Ebright, Y. W. & Gunasekera, A. (1989). Consensus DNA site for the

Escherichia coli catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the *E. coli lac* DNA site. *Nucleic Acids Res* **17**, 10295-10305.

Ebright, R. H. (2000). RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol* **304**, 687-698.

Eiamphungporn, W., Charoenlap, N., Vattanaviboon, P. & Mongkolsuk, S. (2006). *Agrobacterium tumefaciens soxR* is involved in superoxide stress protection and also directly regulates superoxide-inducible expression of itself and a target gene. *J Bacteriol* **188**, 8669-8673.

Epstein, W., Rothman-Denes, L. B. & Hesse, J. (1975). Adenosine 3':5'-cyclic monophosphate as mediator of catabolite repression in *Escherichia coli*. *Proc Natl Acad Sci U S A* **72**, 2300-2304.

Ezraty, B., Aussel, L. & Barras, F. (2005). Methionine sulfoxide reductases in prokaryotes. *Biochim Biophys Acta* **1703**, 221-229.

Gentleman, R. C., Carey, V. J., Bates, D. M. & other authors (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80.

Godde, J. S. & Bickerton, A. (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J Mol Evol* **62**, 718-729.

Gottesman, S. (2003). Proteolysis in bacterial regulatory circuits. *Annu Rev Cell Dev Biol* **19**,

565-587.

Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-308.

Green, J., Scott, C. & Guest, J. R. (2001). Functional versatility in the CRP-FNR superfamily of transcription factors: FNR and FLP. *Adv Microb Physiol* **44**, 1-34.

Gruber, T. M. & Gross, C. A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* **57**, 441-466.

Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* **1**, e60.

Hale, C., Kleppe, K., Terns, R. M. & Terns, M. P. (2008). Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* **14**, 2572-2579.

Harano, Y., Suzuki, I., Maeda, S., Kaneko, T., Tabata, S. & Omata, T. (1997). Identification and nitrogen regulation of the cyanase gene from the cyanobacteria *Synechocystis* sp. strain PCC 6803 and *Synechococcus* sp. strain PCC 7942. *J Bacteriol* **179**, 5744-5750.

Hashimoto, Y., Yano, T., Kuramitsu, S. & Kagamiyama, H. (2001). Disruption of *Thermus thermophilus* genes by homologous recombination using a thermostable kanamycin-resistant marker. *FEBS Lett* **506**, 231-234.

Hecker, M. & Volker, U. (2001). General stress response of *Bacillus subtilis* and other bacteria.

Adv Microb Physiol **44**, 35-91.

Helmann, J. D. (2002). The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol* **46**, 47-110.

Hengge-Aronis, R. (1996). Back to log phase: sigma S as a global regulator in the osmotic control of gene expression in *Escherichia coli*. *Mol Microbiol* **21**, 887-893.

Hengge-Aronis, R. (2002). Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. *Microbiol Mol Biol Rev* **66**, 373-395, table of contents.

Hengge, R. (2008). The two-component network and the general stress sigma factor RpoS (sigma S) in *Escherichia coli*. *Adv Exp Med Biol* **631**, 40-53.

Heyduk, T. & Lee, J. C. (1989). *Escherichia coli* cAMP receptor protein: evidence for three protein conformational states with different promoter binding affinities. *Biochemistry* **28**, 6914-6924.

Holm, L. & Sander, C. (1998). Touring protein fold space with Dali/FSSP. *Nucleic Acids Res* **26**, 316-319.

Horvath, P. & Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167-170.

Hsiao, Y. M., Liao, H. Y., Lee, M. C., Yang, T. C. & Tseng, Y. H. (2005). Clp up-regulates transcription of *engA* gene encoding a virulence factor in *Xanthomonas campestris* by direct

binding to the upstream tandem Clp sites. *FEBS Lett* **579**, 3525-3533.

Imagawa, T., Iino, H., Kanagawa, M., Ebihara, A., Kuramitsu, S. & Tsuge, H. (2008). Crystal structure of the YdjC-family protein TTHB029 from *Thermus thermophilus* HB8: structural relationship with peptidoglycan N-acetylglucosamine deacetylase. *Biochem Biophys Res Commun* **367**, 535-541.

Imlay, J. A. (2002). How oxygen damages microbes: oxygen tolerance and obligate anaerobiosis. *Adv Microb Physiol* **46**, 111-153.

Jansen, R., Embden, J. D., Gaastra, W. & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**, 1565-1575.

Jasiecki, J. & Wegrzyn, G. (2003). Growth-rate dependent RNA polyadenylation in *Escherichia coli*. *EMBO rep* **4**, 172-177.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637.

Kanack, K. J., Runyen-Janecky, L. J., Ferrell, E. P., Suh, S.-J. & West, S. E. H. (2006). Characterization of DNA-binding specificity and analysis of binding sites of the *Pseudomonas aeruginosa* global regulator, Vfr, a homologue of the *Escherichia coli* cAMP receptor protein. *Microbiol* **152**, 3485-3496.

Khorchid, A. & Ikura, M. (2006). Bacterial histidine kinase as signal sensor and transducer. *Int J Biochem Cell Biol* **38**, 307-312.

Kiley, P. J. & Beinert, H. (2003). The role of Fe-S proteins in sensing and regulation in bacteria. *Curr Opin Microbiol* **6**, 181-185.

Kim, D. Y. & Kim, K. K. (2005). Structure and function of HtrA family proteins, the key players in protein quality control. *J Biochem Mol Biol* **38**, 266-274.

Kolb, A., Busby, S., Buc, H., Garges, S. & Adhya, S. (1993). Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem* **62**, 749-795.

Korner, H., Sofia, H. J. & Zumft, W. G. (2003). Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol Rev* **27**, 559-592.

Kuramitsu, K. & Yoshida, A. (1990). Plasma and synovial fluid levels of granulocytal elastase-alpha-1-protease inhibitor complex in patients with rheumatoid arthritis. *Rheumatol Int* **10**, 51-56.

Kyriakidis, D. A. & Tiligada, E. (2009). Signal transduction and adaptive regulation through bacterial two-component systems: the *Escherichia coli* AtoSC paradigm. *Amino Acids* **37**, 443-458.

Larkin, M. A., Blackshields, G., Brown, N. P. & other authors (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

Lawson, C. L., Swigon, D., Murakami, K. S., Darst, S. A., Berman, H. M. & Ebright, R. H. (2004). Catabolite activator protein: DNA binding and transcription activation. *Curr Opin*

Struct Biol **14**, 10-20.

Lee, J. W. & Helmann, J. D. (2006). The PerR transcription factor senses H_2O_2 by metal-catalysed histidine oxidation. *Nature* **440**, 363-367.

Lehmann, C., Lim, K., Chalamasetty, V. R. & other authors (2003). The HI0073/HI0074 protein pair from *Haemophilus influenzae* is a member of a new nucleotidyltransferase family: structure, sequence analyses, and solution studies. *Proteins* **50**, 249-260.

Lehmann, C., Pullalarevu, S., Krajewski, W., Willis, M. A., Galkin, A., Howard, A. & Herzberg, O. (2005). Structure of HI0073 from *Haemophilus influenzae*, the nucleotide-binding domain of a two-protein nucleotidyl transferase. *Proteins* **60**, 807-811.

Leichert, L. I., Scharf, C. & Hecker, M. (2003). Global characterization of disulfide stress in *Bacillus subtilis*. *J Bacteriol* **185**, 1967-1975.

LeMaster, D. M. & Richards, F. M. (1985). 1H-15N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. *Biochemistry* **24**, 7263-7268.

Letek, M., Valbuena, N., Ramos, A., Ordonez, E., Gil, J. A. & Mateos, L. M. (2006). Characterization and use of catabolite-repressed promoters from gluconate genes in *Corynebacterium glutamicum*. *J Bacteriol* **188**, 409-423.

Lu, X. & Zhang, X. (2006). The effect of GeneChip gene definitions on the microarray study of cancers. *Bioessays* **28**, 739-746.

Ludwig, M. L., Metzger, A. L., Pattridge, K. A. & Stallings, W. C. (1991). Manganese superoxide dismutase from *Thermus thermophilus*. A structural model refined at 1.8 A resolution. *J Mol Biol* **219**, 335-358.

Maghnouj, A., Abu-Bakr, A. A., Baumberg, S., Stalon, V. & Vander Wauven, C. (2000). Regulation of anaerobic arginine catabolism in *Bacillus licheniformis* by a protein of the Crp/Fnr family. *FEMS Microbiol Lett* **191**, 227-234.

Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1**, 7.

Makarova, K. S., Wolf, Y. I., van der Oost, J. & Koonin, E. V. (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol Direct* **4**, 29.

Marchler-Bauer, A., Panchenko, A. R., Shoemaker, B. A., Thiessen, P. A., Geer, L. Y. & Bryant, S. H. (2002). CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res* **30**, 281-283.

Marles-Wright, J. & Lewis, R. J. (2007). Stress responses of bacteria. *Curr Opin Struct Biol* **17**, 755-760.

Marraffini, L. A. & Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene

transfer in staphylococci by targeting DNA. *Science* **322**, 1843-1845.

Marraffini, L. A. & Sontheimer, E. J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568-571.

Martin, G. & Keller, W. (2007). RNA-specific ribonucleotidyl transferases. *RNA* **13**, 1834-1849.

McRee, D. E. (1999). XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* **125**, 156-165.

Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**, 174-182.

Nakagawa, N., Sugahara, M., Masui, R., Kato, R., Fukuyama, K. & Kuramitsu, S. (1999). Crystal structure of *Thermus thermophilus* HB8 UvrB protein, a key enzyme of nucleotide excision repair. *J Biochem* **126**, 986-990.

Nakunst, D., Larisch, C., Hüser, A. T., Tauch, A., Pühler, A. & Kalinowski, J. (2007). The extracytoplasmic function-type sigma factor SigM of *Corynebacterium glutamicum* ATCC 13032 is involved in transcription of disulfide stress-related genes. *J Bacteriol* **189**, 4696-4707.

Nandi, D. L., Horowitz, P. M. & Westley, J. (2000). Rhodanese as a thioredoxin oxidase. *Int J Biochem Cell Biol* **32**, 465-473.

Naryshkina, T., Liu, J., Florens, L. & other authors (2006). *Thermus thermophilus* bacteriophage phiYS40 genome and proteomic characterization of virions. *J Mol Biol* **364**, 667-677.

Nyström, T. (2004). Stationary-phase physiology. *Ann Rev Microbiol* **58**, 161-181.

Ooga, T., Ohashi, Y., Kuramitsu, S., Koyama, Y., Tomita, M., Soga, T. & Masui, R. (2009). Degradation of ppGpp by nudix pyrophosphatase modulates the transition of growth phase in the bacterium *Thermus thermophilus*. *J Biol Chem* **284**, 15549-15556.

Oshima, T., and K. Imabori (1974). Description of *Thermus thermophilus* (Yoshida and Oshima) com. nov., a non-sporulating thermophilic bacterium from a Japanese thermal spa. *Int J Syst Bacteriol* **24**, 102-112.

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods Enzymol*, pp. 307-326.

Park, H. J., Reiser, C. O., Kondruweit, S., Erdmann, H., Schmid, R. D. & Sprinzl, M. (1992). Purification and characterization of a NADH oxidase from the thermophile *Thermus thermophilus* HB8. *Eur J Biochem* **205**, 881-885.

Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. E. & Berman, H. M. (1996). Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface. *J Mol Biol* **260**, 395-408.

Passner, J. M. & Steitz, T. A. (1997). The structure of a CAP-DNA complex having two cAMP

molecules bound to each monomer. *Proc Natl Acad Sci U S A* **94**, 2843-2847.

Passner, J. M., Schultz, S. C. & Steitz, T. A. (2000). Modeling the cAMP-induced allosteric transition using the crystal structure of CAP-cAMP at 2.1 A resolution. *J Mol Biol* **304**, 847-859.

Pepper, S. D., Saunders, E. K., Edwards, L. E., Wilson, C. L. & Miller, C. J. (2007). The utility of MAS5 expression summary and detection call algorithms. *BMC Bioinformatics* **8**, 273.

Perrakis, A., Harkiolaki, M., Wilson, K. S. & Lamzin, V. S. (2001). ARP/wARP and molecular replacement. *Acta Crystallogr D Biol Crystallogr* **57**, 1445-1450.

Peterson, J., Fee, J. A. & Day, E. P. (1991). Magnetization of manganese superoxide dismutase from *Thermus thermophilus*. *Biochim Biophys Acta* **1079**, 161-168.

Pomposiello, P. J. & Demple, B. (2001). Redox-operated genetic switches: the SoxR and OxyR transcription factors. *Trends Biotechnol* **19**, 109-114.

Pourcel, C., Salvignol, G. & Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653-663.

Raivio, T. L. & Silhavy, T. J. (2001). Periplasmic stress and ECF sigma factors. *Ann Rev Microbiol* **55**, 591-624.

Reeve, C. A., Bockman, A. T. & Matin, A. (1984). Role of protein degradation in the survival

of carbon-starved *Escherichia coli* and *Salmonella typhimurium*. *J Bacteriol* **157**, 758-763.

Rehse, P. H., Ohshima, N., Nodake, Y. & Tahirov, T. H. (2004). Crystallographic structure and biochemical analysis of the *Thermus thermophilus* osmotically inducible protein C. *J Mol Biol* **338**, 959-968.

Rock, C. O. & Cronan, J. E. (1996). *Escherichia coli* as a model for the regulation of dissociable (type II) fatty acid biosynthesis. *Biochim Biophys Acta* **1302**, 1-16.

Sakaki, Y. & Oshima, T. (1975). Isolation and characterization of a bacteriophage infectious to an extreme thermophile, *Thermus thermophilus* HB8. *J Virol* **15**, 1449-1453.

Sakamoto, K., Agari, Y., Yokoyama, S., Kuramitsu, S. & Shinkai, A. (2008). Functional identification of an anti-sigmaE factor from *Thermus thermophilus* HB8. *Gene* **423**, 153-159.

Sakamoto, K., Agari, Y., Agari, K., Yokoyama, S., Kuramitsu, S. & Shinkai, A. (2009). X-ray crystal structure of a CRISPR-associated RAMP superfamily protein, Cmr5, from *Thermus thermophilus* HB8. *Proteins* **75**, 528-532.

Sakamoto, K., Agari, Y., Agari, K., Kuramitsu, S. & Shinkai, A. (2010). Structural and functional characterization of the transcriptional repressor CsoR from *Thermus thermophilus* HB8. *Microbiology* **156**, 1993-2005.

Sandberg, R. & Larsson, O. (2007). Improved precision and accuracy for microarrays using updated probe set definitions. *BMC Bioinformatics* **8**, 48.

Sanger, F., Nicklen, S. & Coulson, A. R. (1992). DNA sequencing with chain-terminating

inhibitors. 1977. *Biotechnology* **24**, 104-108.

Santos, J. M., Freire, P., Mesquita, F. S., Mika, F., Hengge, R. & Arraiano, C. M. (2006). Poly(A)-polymerase I links transcription with mRNA degradation via σ^S proteolysis. *Mol Microbiol* **60**, 177-188.

Schellhorn, H. E. (1995). Regulation of hydroperoxidase (catalase) expression in *Escherichia coli*. *FEMS Microbiol Lett* **131**, 113-119.

Schneider, T. D. & Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* **18**, 6097-6100.

Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science* **253**, 1001-1007.

Scott, S. P. & Jarjous, S. (2005). Proposed structural mechanism of *Escherichia coli* cAMP receptor protein cAMP-dependent proteolytic cleavage protection and selective and nonselective DNA binding. *Biochemistry* **44**, 8730-8748.

Sevostyanova, A., Djordjevic, M., Kuznedelov, K., Naryshkina, T., Gelfand, M. S., Severinov, K. & Minakhin, L. (2007). Temporal regulation of viral transcription during development of *Thermus thermophilus* bacteriophage phiYS40. *J Mol Biol* **366**, 420-435.

Shah, S. A., Hansen, N. R. & Garrett, R. A. (2009). Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. *Biochem Soc Trans* **37**, 23-28.

Sharma, H., Yu, S., Kong, J., Wang, J. & Steitz, T. A. (2009). Structure of apo-CAP reveals that large conformational changes are necessary for DNA binding. *Proc Natl Acad Sci U S A* **106**, 16604-16609.

Shimada, A., Masui, R., Nakagawa, N., Takahata, Y., Kim, K., Kuramitsu, S. & Fukui, K. (2010). A novel single-stranded DNA-specific 3'-5' exonuclease, *Thermus thermophilus* exonuclease I, is involved in several DNA repair pathways. *Nucleic Acids Res* **38**, 5692-5705.

Shinkai, A., Patel, P. H. & Loeb, L. A. (2001). The conserved active site motif A of *Escherichia coli* DNA polymerase I is highly mutable. *J Biol Chem* **276**, 18836-18842.

Shinkai, A., Kira, S., Nakagawa, N., Kashihara, A., Kuramitsu, S. & Yokoyama, S. (2007a). Transcription Activation Mediated by a Cyclic AMP Receptor Protein from *Thermus thermophilus* HB8. *J Bacteriol* **189**, 3891-3901.

Shinkai, A., Ohbayashi, N., Terada, T., Shirouzu, M., Kuramitsu, S. & Yokoyama, S. (2007b). Identification of promoters recognized by RNA polymerase-σ^E holoenzyme from *Thermus thermophilus* HB8. *J Bacteriol* **189**, 8758-8764.

Sorek, R., Kunin, V. & Hugenholtz, P. (2008). CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Rev Microbiol* **6**, 181-186.

Stock, A. M., Robinson, V. L. & Goudreau, P. N. (2000). Two-component signal transduction. *Annu Rev Biochem* **69**, 183-215.

Storey, J. D. (2002). A direct approach to false discovery rates. *J R Stat Soc B* **64**, 479-498.

Storz, G. & Tartaglia, L. A. (1992). OxyR: a regulator of antioxidant genes. *J Nutr* **122**, 627-630.

Storz, G. & Imlay, J. A. (1999). Oxidative stress. *Curr Opin Microbiol* **2**, 188-194.

Sweetser, D., Nonet, M. & Young, R. A. (1987). Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc Natl Acad Sci U S A* **84**, 1192-1196.

Tanigawa, R., Shirokane, M., Maeda Si, S., Omata, T., Tanaka, K. & Takahashi, H. (2002). Transcriptional activation of NtcA-dependent promoters of Synechococcus sp. PCC 7942 by 2-oxoglutarate in vitro. *Proc Natl Acad Sci U S A* **99**, 4251-4255.

Tatusov, R. L., Fedorova, N. D., Jackson, J. D. & other authors (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41.

Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* **55**, 849-861.

Touati, D. (2000). Sensing and protecting against superoxide stress in *Escherichia coli*--how many ways are there to trigger *soxRS* response? *Redox Rep* **5**, 287-293.

Tsilibaris, V., Maenhaut-Michel, G. & Van Melderen, L. (2006). Biological roles of the Lon ATP-dependent protease. *Res Microbiol* **157**, 701-713.

Vázquez-Bermúdez, M. F., Herrero, A. & Flores, E. (2002). 2-Oxoglutarate increases the binding affinity of the NtcA (nitrogen control) transcription factor for the *Synechococcus glnA* promoter. *FEBS Lett* **512**, 71-74.

van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M. & Brouns, S. J. (2009). CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci* **34**, 401-407.

Vassylyeva, M. N., Lee, J., Sekine, S. I. & other authors (2002). Purification, crystallization and initial crystallographic analysis of RNA polymerase holoenzyme from *Thermus thermophilus*. *Acta Crystallogr D Biol Crystallogr* **58**, 1497-1500.

Vicente, M., Chater, K. F. & De Lorenzo, V. (1999). Bacterial transcription factors involved in global regulation. *Mol Microbiol* **33**, 8-17.

Wöhlkonig, A., Stalon, V. & Vander Wauven, C. (2004). Purification of ArcR, an oxidation-sensitive regulatory protein from *Bacillus licheniformis*. *Protein Expr Purif* **37**, 32-38.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. (2008). Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**, 209-213.

Watanabe, S., Muramatsu, T., Ao, H., Hirayama, Y., Takahashi, K., Tanokura, M. & Kuchino, Y. (1999). Molecular cloning of the Lon protease gene from *Thermus thermophilus* HB8 and characterization of its gene product. *Eur J Biochem* **266**, 811-819.

Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W. & Doudna, J. A. (2009). Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* **17**, 904-912.

Wnendt, S., Hartmann, R. K., Ulbrich, N. & Erdmann, V. A. (1990). Isolation and physical properties of the DNA-directed RNA polymerase from *Thermus thermophilus* HB8. *Eur J Biochem* **191**, 467-472.

Yang, J., Bitoun, J. P. & Ding, H. (2006). Interplay of IscA and IscU in biogenesis of iron-sulfur clusters. *J Biol Chem* **281**, 27956-27963.

Yokoyama, S., Hirota, H., Kigawa, T. & other authors (2000a). Structural genomics projects in Japan. *Nat Struct Biol* **7 Suppl**, 943-945.

Yokoyama, S., Matsuo, Y., Hirota, H. & other authors (2000b). Structural genomics projects in Japan. *Prog Biophys Mol Biol* **73**, 363-376.

Yu, M. X., Slater, M. R. & Ackermann, H. W. (2006). Isolation and characterization of *Thermus* bacteriophages. *Arch Virol* **151**, 663-679.

Zheng, L., Cash, V. L., Flint, D. H. & Dean, D. R. (1998). Assembly of iron-sulfur clusters. Identification of an *iscSUA-hscBA-fdx* gene cluster from *Azotobacter vinelandii*. *J Biol Chem* **273**, 13264-13272.

Acknowledgements

I would like to express my great appreciation to Professor Seiki Kuramitsu and Dr. Akeo Shinkai (RIKEN) for their guidance and many valuable discussions. I also would like to express my great appreciation to Professors Satoru Kawamura, Katsuyuki Tanizawa, and Akihiko Ogura for their critical advice for this study. I am grateful to Ms. Keiko Sakamoto for help with the RT-PCR and the phage infection analyses, and Mses. Aiko Kashihara, Emi Ishido-Nakai, and Miwa Ohmori for their help with the DNA microarray dataset collection. I would also like to thank Ms. Masami Nishida for help with the purification of the Se-SdrP protein, Mses. Kazuko Agari and Aimi Osaki for their help with the crystallization of the SdrP protein, and Ms. Kazuko Agari for help with the template preparation for the *in vitro* transcription assays. Finally, I would like to thank my colleagues in the Kuramitsu laboratory, the SR System Biology Research Group of RIKEN SPring-8 Center and my family for their kind help and encouragement.

List of publications

The contents of this doctoral thesis are mainly included in the following articles.

Agari, Y., Kashihara, A., Yokoyama, S., Kuramitsu, S. & Shinkai, A. (2008). Global gene expression mediated by *Thermus thermophilus* SdrP, a CRP/FNR family transcriptional regulator. *Mol Microbiol* **70**, 60-75.

Agari, Y., Kuramitsu, S. & Shinkai, A. (2010). Identification of novel genes regulated by the oxidative stress-responsive transcriptional activator SdrP in *Thermus thermophilus* HB8. *FEMS microbiol lett* **313**, 127-134.

Agari, Y., Sakamoto, K., Tamakoshi, M., Oshima, T., Kuramitsu, S. & Shinkai, A. (2010). Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. *J Mol Biol* **395**, 270-281.

The other articles are listed below.

Agari, Y., Agari, K., Sakamoto, K., Kuramitsu, S. & Shinkai, A. (2011). TetR family transcriptional repressor *Thermus thermophilus* FadR controls fatty acid degradation. *Microbiology*, in press.

Agari, Y., Yokoyama, S., Kuramitsu, S. & Shinkai, A. (2008). X-ray crystal structure of a CRISPR-associated protein, Cse2, from *Thermus thermophilus* HB8. *Proteins* **73**, 1063-1067.

Agari, Y., Sato, S., Wakamatsu, T., Bessho, Y., Ebihara, A., Yokoyama, S., Kuramitsu, S. & Shinkai, A. (2008). X-ray crystal structure of a hypothetical Sua5 protein from *Sulfolobus*

tokodaii strain 7. *Proteins* **70**, 1108-1111.

Sakamoto, K., Agari, Y., Agari, K., Kuramitsu, S. & Shinkai, A. (2010). Structural and functional characterization of the transcriptional repressor CsoR from *Thermus thermophilus* HB8. *Microbiology* **156**, 1993-2005.

Sakamoto, K., Agari, Y., Agari, K., Yokoyama, S., Kuramitsu, S. & Shinkai, A. (2009). X-ray crystal structure of a CRISPR-associated RAMP superfamily protein, Cmr5, from *Thermus thermophilus* HB8. *Proteins* **75**, 528-532.

Sakamoto, K., Agari, Y., Yokoyama, S., Kuramitsu, S. & Shinkai, A. (2008). Functional identification of an anti-σ^E factor from *Thermus thermophilus* HB8. *Gene* **423**, 153-159.

Padmanabhan, B., Strange, R. W., Antonyuk, S. V., Ellis, M. J., Hasnain, S. S., Iino, H., Agari, Y., Bessho, Y. & Yokoyama, S. (2009). Structure of dihydrodipicolinate synthase from *Methanocaldococcus jannaschii*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **65**, 1222-1226.

Kitamura, Y., Ebihara, A., Agari, Y., Shinkai, A., Hirotsu, K. & Kuramitsu, S. (2009). Structure of D-alanine-D-alanine ligase from *Thermus thermophilus* HB8: cumulative conformational change and enzyme-ligand interactions. *Acta Crystallogr D Biol Crystallogr* **65**, 1098-1106.

Ragunathan, P., Kumarevel, T., Agari, Y., Shinkai, A., Kuramitsu, S., Yokoyama, S. & Ponnuraj, K. (2008). Crystal structure of ST2348, a CBS domain protein, from

hyperthermophilic archaeon *Sulfolobus tokodaii*. *Biochem biophys res commun* **375**, 124-128.

Rangarajan, S., Jeyakanthan, J., Mridula, P., Sakamoto, K., Kitamura, Y., Agari, Y., Shinkai, A., Ebihara, A., Kuramitsu, S., Yokoyama, S. & Sekar, K. (2008). Crystallization and preliminary crystallographic studies of L30e, a ribosomal protein from *Methanocaldo-coccus jannaschii* (MJ1044). *Acta Crystallogr Sect F Struct Biol Cryst Commun* **64**, 102-104.

Iino, H., Naitow, H., Nakamura, Y. Nakamura, Y., Nakagawa, N., Agari, Y., Kanagawa, M., Ebihara, A., Shinkai, A., Sugahara, M., Miyano, M., Kamiya, N., Yokoyama, S., Hirotsu, K. & Kuramitsu, S. (2008). Crystallization screening test for the whole-cell project on *Thermus thermophilus* HB8. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **64**, 487-491.

Yamamura, A., Ohtsuka, J., Kubota, K., Agari, Y., Ebihara, A., Nakagawa, N., Nagata, K. & Tanokura, M. (2008). Crystal structure of TTHA1429, a novel metallo-beta-lactamase superfamily protein from *Thermus thermophilus* HB8. *Proteins* **73**, 1053-1057.

Goto, M., Agari, Y., Omi, R., Miyahara, I. & Hirotsu, K. (2003). Expression, purification and preliminary X-ray characterization of N-acetyl-gamma-glutamyl-phosphate reductase from *Thermus thermophilus* HB8. *Acta Crystallogr D Biol Crystallogr* **59**, 356-358.

Ohtani, N., Nakagawa, N., Hoseki, J. Ebihara, A., Satoh, S., Agari, Y., Kobayashi, S., Agari, K., Maoka, N., Masui, R., Miki, K., Yokoyama, S. & Kuramitsu, S. (2002). [An exhaustive overproduction of bacterial proteins]. *Tanpakushitsu kakusan koso* **47**, 1009-1013.