# An archaeal transcription factor EnfR with a novel 'eighth note' fold controls hydrogen production of a hyperthermophilic archaeon *Thermococcus onnurineus* NA1

Da-Woon Bae<sup>1,†</sup>, Seong Hyuk Lee<sup>2,†</sup>, Ji Hye Park<sup>3</sup>, Se-Young Son<sup>1</sup>, Yuxi Lin<sup>4</sup>, Jung Hyen Lee<sup>3</sup>, Bo-Ram Jang<sup>5</sup>, Kyu-Ho Lee<sup>5</sup>, Young-Ho Lee<sup>4,6,7,8</sup>, Hyun Sook Lee<sup>2,9</sup>, Sung Gyun Kang<sup>2,9</sup>, Byoung Sik Kim<sup>3</sup> and Sun-Shin Cha<sup>91,\*</sup>

<sup>1</sup>Department of Chemistry & Nanoscience, Ewha Womans University, Seoul 03760, Republic of Korea, <sup>2</sup>Marine Biotechnology Research Center, Korea Institute of Ocean Science and Technology, Busan, South Korea,
<sup>3</sup>Department of Food Science and Biotechnology, Ewha Womans University, Seoul 03760, Republic of Korea,
<sup>4</sup>Research Center for Bioconvergence Analysis, Korea Basic Science Institute (KBSI), Cheongju, Chungbuk 28119, Republic of Korea, <sup>5</sup>Department of Life Science, Sogang University, 35 Baekbeom-Ro, Mapo-Gu, Seoul, South Korea., <sup>6</sup>Bio-Analytical Science, University of Science and Technology, Daejeon 34113, Republic of Korea,
<sup>7</sup>Department of Systems Biotechnology, Chung-Ang University, Anseong, Gyeonggi 17546, Republic of Korea,
<sup>8</sup>Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai, Miyagi 980-8578, Japan and
<sup>9</sup>Department of Marine Biotechnology, KIOST School, University of Science and Technology, Daejeon, South Korea

Received December 27, 2022; Revised July 13, 2023; Editorial Decision August 07, 2023; Accepted August 14, 2023

# ABSTRACT

Thermococcus onnurineus NA1, a hyperthermophilic carboxydotrophic archaeon, produces H<sub>2</sub> through CO oxidation catalyzed by proteins encoded in a carbon monoxide dehydrogenase (CODH) gene cluster. TON\_1525 with a DNA-binding helixturn-helix (HTH) motif is a putative repressor regulating the transcriptional expression of the codh gene cluster. The T55I mutation in TON\_1525 led to enhanced H<sub>2</sub> production accompanied by the increased expression of genes in the codh cluster. Here, TON\_1525 was demonstrated to be a dimer. Monomeric TON\_1525 adopts a novel 'eighth note' symbol-like fold (referred to as 'eighth note' fold regulator, EnfR), and the dimerization mode of EnfR is unique in that it has no resemblance to structures in the Protein Data Bank. According to footprinting and gel shift assays, dimeric EnfR binds to a 36-bp pseudo-palindromic inverted repeat in the promoter region of the *codh* gene cluster, which is supported by an in silico EnfR/DNA complex model and mutational studies revealing the implication of N-terminal loops as well as HTH motifs in DNA recognition. The DNA-binding affinity of the T55I mutant was lowered

by ~15-fold, for which the conformational change of N-terminal loops is responsible. In addition, transcriptome analysis suggested that EnfR could regulate diverse metabolic processes besides  $H_2$  production.

# **GRAPHICAL ABSTRACT**



<sup>\*</sup>To whom correspondence should be addressed. Tel: +82 2 3277 3858; Fax: +82 2 3277 4546; Email: chajung@ewha.ac.kr <sup>†</sup>The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Fossil fuels such as coal, oil and natural gas have been major energy sources responsible for almost 80% of a total energy supply for decades (1,2). However, due to air pollution and global warming caused by their combustion, fossil fuel consumption is declining at a global level and the demand for alternative renewable energy sources is soaring (3). According to a roadmap to 2050 reported by the International Renewable Energy Agency (IRENA), the percentage share of renewable energy would be reached 65% in 2050, whereas the use of fossil fuels would fall to one-third of the present levels (4). Hydrogen is an attractive candidate as a future fuel for several reasons. It is a non-toxic, clean and efficient source in that the major byproduct from hydrogen combustion is pure water, and its energy content per unit mass is 3-times higher than traditional fossil fuels (5). Therefore, the global demand for hydrogen energy is getting higher, and the development of hydrogen-fueling infrastructures for production, storage and utilization is a great research challenge (6-8).

Almost 96% of hydrogen is produced using natural gas (48%), petroleum (30%) and coal (18%) (9). Besides H<sub>2</sub> production depending on nonrenewable fossil sources, there are ways to produce H<sub>2</sub> from renewable feedstocks such as wind, solar and biomass (10). Microbial  $H_2$  production is also promising since it is an environment-friendly and less energy-intensive way to produce H<sub>2</sub>. It is also notable that microbes can exploit industrial wastes including carbon monoxide (CO) or organic matter to produce  $H_2$  (11–13). In particular, biohydrogen production using thermophiles at high temperatures ( $\geq 60^{\circ}$ C) has numerous advantages such as high metabolic activity leading to enhanced product formation rates (14), less contamination by  $H_2$ -consuming microorganisms (15), no intensive cooling necessary during fermentation, decreased density, surface tension, and viscosity of culture broth, etc. (16). Thermococcales, Thermotogales, Desulfurococcales and Clostridium species have been reported to produce H<sub>2</sub> from organic substrates at high temperatures (17-19).

Thermococcus onnurineus NA1, isolated from a deep-sea hydrothermal vent, is a sulfur-reducing hyperthermophilic carboxydotrophic archaeon (20). It requires elemental sulfur as a terminal electron acceptor for heterotrophic growth on peptides or an amino acid mixture and exhibits optimum growth at 80°C and pH 8.5. Interestingly, T. onnurineus NA1 grows under diverse substrates including CO and formate to produce  $H_2$  (21). The  $H_2$  productivity of T. onnurineus NA1 is higher than that of other hyperthermophilic archaea such as Pyrococcus furiosus and Desulfurococcus amylolyticus (18,21,22). In the case of the CO-utilization, the unique codh gene cluster in T. onnurineus NA1, which encodes CO dehydrogenase, hydrogenase and Na<sup>+</sup>/H<sup>+</sup> antiporter, catalyzes the anaerobic oxidation of CO to CO<sub>2</sub> through the water-gas shift reaction:  $\text{CO} + \text{H}_2\text{O} \rightarrow \text{H}_2 + \text{CO}_2 (\Delta G^{\circ\prime} = -20 \text{ kJ/mol}) (23).$ 

In a previous study, an adaptive evolution was carried out through the serial transfer of *T. onnurineus* NA1 into CO medium to enhance its  $H_2$  production (24). After the 156time serial transfer, cell density, CO consumption rate, and  $H_2$  production rate increased by 2.8-, 6.5- and 5.9-fold, respectively. Although mutations occurred in gene sequences of ten proteins after the serial transfer (24), the CO consumption rate and H<sub>2</sub> production rate of a mutant strain MC11 harboring the Thr  $\rightarrow$  Ile replacement at position 55 in the TON\_1525 protein were comparable to those of the 156-serial-transferred strain (156T). Consistently, the expression of proteins essential for CO-dependent H<sub>2</sub> production was increased in both 156T and MC11 strains compared to the wild-type (24).

Sequence analysis revealed that TON\_1525 is a Tfx DNA-binding protein. According to NCBI's Gene resources, a total of 288 microorganisms have Tfx DNAbinding proteins, most of which ( $\sim 98\%$ ) are distributed in archaea, including Thermococcales, Methanococcales, and Halobacteriales. Almost all archaeal Tfx DNA-binding proteins have a basic DNA-binding helix-turn-helix (HTH) motif (p*I* range = 8.8-11.0) at their N-terminal region and an acidic C-terminal domain (pI range = 4.6-6.8) in common. The amino acid sequence of the HTH motif resembles that of the bacterial RNA polymerase (RNAP)  $\sigma^{70}$  region4 domain that recognizes specific transcription-initiation sites in the promoter region and recruits RNAP to start transcription (25). The presence of the DNA-recognizing HTH motif, together with experimental data that two Tfx family members (TON\_1525 and the MTH0916 protein from Methanobacterium thermoautotrophicum) bind to promoter regions (24,26), strongly suggests that the Tfx family members are transcription regulators. Nevertheless, there is no report about the basic characteristics of Tfx family members such as the molecular shape, the oligomerization state, the target DNA sequence, and the DNA-recognition mode. In this study, we revealed that TON\_1525 adopts a fold resembling an eighth note symbol and the unique X-shaped dimeric structure of EnfR has no structural homolog. In addition, the target DNA sequence was determined in the promoter region of the codh gene cluster, and the DNAbinding mode of EnfR was elucidated based on structural and mutational analyses. We also provide a possibility that it may be involved in the transcriptional regulation of diverse metabolic processes in T. onnurineus NA1. Consequently, our study will play as a platform for future research about the Tfx family members and for the understanding of how the energy metabolism of a hyperthermophilic archaeon T. onnurineus NA1 is controlled at the transcription level.

#### MATERIALS AND METHODS

#### **Protein preparation**

The *T. onnurineus* NA1 TON\_1525 gene encoding residues 1–146 was synthesized to have the in-frame non-cleavable C-terminal His6-tag and inserted at the *NdeI* and *NotI* sites of the expression vector pET24a(+) (Novagen, USA). The T55I substitution was generated by site-directed mutagenesis using the wild-type gene as a template. The wild-type and mutant construct were transformed into *Escherichia coli* Rosetta (DE3), respectively. The two transformed cells were grown to an optical density values at 600 nm (OD<sub>600</sub>) of ~0.6 in Luria-Bertani medium at 37°C, and the expression of the wild-type and mutants were induced by 0.5 mM isopropyl  $\beta$ -D-thiogalactopyranoside. After 4 h induction at 37°C, the cells were harvested, resuspended in a 10 mM

Tris (pH 7.4), 500 mM NaCl and 5 mM  $\beta$ -mercaptoethanol buffer, and disrupted by sonication. The crude lysate was centrifugated at 10,000 × g for 30 min and the resulting supernatant was then boiled at 75°C for 10 min. The lysate was centrifugated again at 10,000 × g for 1 h. The resulting supernatant was loaded onto a nickel-nitrilotriacetic acid (Ni-NTA) column (GE Healthcare, USA). The protein fraction eluted from the Ni-NTA column was concentrated and loaded onto a Superdex 75 HR 16/600 column (GE Healthcare, USA). The elute from gel-filtration was concentrated to ~10 mg/ml in a 20 mM Tris buffer (pH 7.4), 500 mM NaCl and 2 mM DTT for crystallization.

# Crystallization, X-ray diffraction experiment and structure determination

The microbatch crystallization method was employed to grow crystals under oil (27). Crystals of the wild-type TON\_1525 were obtained at 22°C by mixing 1 µl of protein solution with an equivalent volume of a mother liquor consisting of 30% polyethylene glycol (PEG) 400, 100 mM HEPES (pH 7.5), and 200 mM NaCl. The T55I mutant was crystallized in the same way with a mother liquor of 30% PEG 400, 100 mM HEPES (pH 7.5), and 200 mM ammonium sulfate. A 2.8 Å resolution data set of the wild-type and a 3.4 Å resolution data set of the T55I mutant were collected at beamline 17A of Photon Factory, Tsukuba, Japan. Both data sets were integrated and scaled with XDS (28). Crystals of the wild-type belonged to the space group  $P4_{3}2_{1}2$  with cell parameters a = b = 102.7 Å and c = 90.4Å corresponding to two monomers in an asymmetric unit. Crystals of the T55I mutant, which belonged to the space group  $P4_32_12$  with cell parameters a = b = 123.2 Å and  $c = \hat{8}2.7$  Å, contain two monomers in an asymmetric unit (Supplementary Table S1).

To solve the structures of the wild-type and the T55I mutant by molecular replacement (MR), we used the MTH0916 structure (PDB code:1NR3) as a search model because MTH0916 is the only Tfx family member whose structure is available. However, our MR trials using programs MOLREP and PHASER were unsuccessful. TON\_1525 has no methionine residue except for the Nterminal initiation methionine that is usually disordered in proteins, which indicates that TON\_1525 is not adequate for incorporating seleno-methionines for de novo phasing. Therefore, a leucine residue at position 83 was replaced by methionine together with another C135S substitution in the T55I mutant. Since a protein band corresponding to the dimeric size of TON\_1525 was observed on non-reducing SDS-PAGE during protein purification, we mutated Cys135 to serine to prevent intermolecular disulfide bonds. After the introduction of the Cys  $\rightarrow$  Ser replacement, the band of the dimeric size disappeared. For the seleno-methionine labeling of the triple mutant (T55I/L83M/C135S), the methionine auxotroph E. coli B834 (DE3) (Novagen) strain was used as a host. The triple mutant harboring a selenomethionine at position 83 was purified in the same way as the wild-type and was concentrated to  $\sim 10 \text{ mg/ml}$  for crystallization.

Crystals of the triple mutant were obtained in the same way as the wild-type with a mother liquor consisting of 16% PEG 400, 100 mM Tris (pH 8.6) and 200 mM ammonium sulfate. A single-wavelength (0.9796Å) anomalous diffraction (SAD) data set was collected to the resolution of 3.00 Å at beamline 5C of Pohang Accelerator Laboratory and was processed by HKL2000 (29). The experimental SAD phasing and the initial model building were performed with *PHENIX* (30). The initial model of the triple mutant was subjected to the automated model building of BUCCANEER (31) and then further refined to  $R/R_{\text{free}}$  of 0.251/0.267 by using COOT (32) and PHENIX. The crystal structures of the wild-type and the T55I mutant were determined with the triple mutant structure as a search model. MR solutions from PHASER was manually manipulated by COOT and refined by PHENIX. Several rounds of refinements and manual refitting gave rise to final models of the wild-type  $(R/R_{\text{free}} = 0.222/0.244)$  and the T55I mutant  $(R/R_{\rm free} = 0.236/0.275)$ . The final model of the wild-type contains residues 4~142 of chain A and residues 4~139 of chain B while that of the T55I mutant consists of residues 4-144 of chain A and residues 2-139 of chain B. The Ramachandran plots indicate 97.1% (the wild-type) and 91.6% (the T55I mutant) of non-glycine residues are in the most favored regions, and all others are in the additionally allowed regions.

#### Analytical ultracentrifugation

Sedimentation velocity of wild-type EnfR and all the mutants were calculated by analytical ultracentrifugation at a wavelength of 280 nm using a ProteomeLab XL-A centrifuge (Beckman Coulter, Brea, CA, USA) equipped with an AN-60 Ti rotor at 20°C. Data from sedimentation at 42,000 rpm were collected at 5-min intervals, giving a total of 130 scans. Sedimentation velocity data were analyzed using *SEDFIT*. The density and viscosity of the buffer solution were determined at 20°C to be 1.019 g/ml and 0.010503 P, respectively. The partial specific volume was calculated as 0.7454–0.7466 ml/g at 20°C.

#### **DNase I protection assay**

For DNase I protection assay, a DNA probe of the 150-bp promoter region of the *codh* gene cluster (24) was amplified by PCR using 6-carboxyfluorescein (6-FAM)-labeled 1017\_150\_F and unlabeled 1017\_150\_R as primers (Supplementary Table S2). The labeled DNA probe (200 ng) was then incubated with the purified EnfR protein for 10 min at 80°C in a 20 µl reaction mixture containing  $1 \times$  binding buffer (20 mM Tris–HCl pH 7.5, 200 mM KCl, 1 mM EDTA, and 5% glycerol). DNaseI digestion of the protein-DNA complex followed the procedures described previously (33). The resulting DNA fragments were precipitated with ethanol, eluted in nuclease-free water and then analyzed using an ABI 3730xl DNA analyzer (Applied Biosystems, Foster City, CA) with Peak Scanner<sup>TM</sup> Software ver. 1.0 (Applied Biosystems).

#### Molecular docking

An *in silico* model of EnfR in complex with the 36-bp target DNA was built using the *HADDOCK* v.2.4 server. For molecular docking, we defined the  $\alpha$ 2-turn- $\alpha$ 3 region (residues 24–51) of EnfR as active residues that were predicted to be involved in DNA-binding, and the first three N-terminal residues that are conformationally different from that of the T55I mutant structure were also selected as active residues were selected based on the sequence of 5'-AATCTTTTGTTTACATT-3' which was determined as a minimal binding region in the DNase I protection assay. Passive residues were automatically selected by the server. Among the generated models, the top-ranked model was selected to get insights into the DNA-binding mode of EnfR.

#### Electrophoretic mobility shift assay (EMSA)

All protein mutants were generated by site-directed mutagenesis using the wild-type TON\_1525 (EnfR) gene as a template (Cosmogenetech, Republic of Korea). Mutant proteins were prepared by the same procedure as described in Protein Preparation. To analyze interactions between EnfR and the target DNA sequence, the 36-bp 5'-Cvanine 5 (Cv5)-labeled (5 nM) probes were used with different protein concentrations. For verification of EnfR-binding to the promoter regions of differentially expressed genes (DEGs), DNA probes (10 nM) of TON\_1582, TON\_0537 and TON\_1563 genes were amplified by PCR using 5'-Cy5labeled forward primers (1582\_F, 0537/8\_F and 1563\_F) and unlabeled reverse primers (1582\_R, 0537/8\_R and 1563\_R) (Supplementary Table S4). Proteins were incubated with the probes at 37°C for 20 min in a buffer containing 10 mM Tris (pH 7.4), 200 mM KCl, 2 mM DTT, 0.2 mM EDTA. After incubation, the reaction mixtures were separated by electrophoresis using a 6% polyacrylamide gel. The gels were analyzed with ChemiDoc MP imaging system (Bio-Rad). The band intensities were calculated with the program ImageJ and used to compare the DNA-binding affinities of the wild-type and the T55I mutant.

#### Strains, culture conditions and analytical methods

*T. onnurineus* NA1 (KCTC 10859) was routinely cultured in modified medium 1 (MM1) supplemented with 1 bar of 100% CO (MM1 + CO) in serum bottles under anaerobic conditions at 80°C (34). An MC11 mutant strain harboring the amino acid substitution, T55I, in EnfR was previously constructed (24) and used in this study. For the cultivation of the wild-type and MC11 strains in the bioreactor, 100% CO was supplemented at a flow rate of 120 ml/min and the pH was controlled at 6.1–6.2 using 5 N NaOH containing 3.5% NaCl. Cell growth was monitored by OD<sub>600</sub>, and H<sub>2</sub> production was analyzed as previously described (35).

#### Transcriptome analysis

The wild-type and MC11 strains were grown to the exponential growth phase in MM1 + CO medium and three independent cultures were employed for RNA extraction. Total RNA was extracted from the cells using Trizol reagent

(Invitrogen, Carlsbad, CA) as previously described (24). The quality and quantity of the RNAs were assessed as RNA integrity number (RIN) (36) and RNA electropherograms produced with an Agilent 2100 Bioanalyzer (Palo Alto, USA) according to the manufacturer's instructions.  $10 \mu g$  of RNA from samples with RIN values ranging from 8.4 to 9.7 for the wild-type and MC11 strains was used for further experiments. Sequencing library generation and sequencing using a HiSeq 2500 (Illumina, San Diego, USA) were performed by CJ bioscience (Seoul, Republic of Korea). Quality-filtered reads were mapped against a reference genome sequence (BioProject accession number PR-JNA59043) (https://www.ncbi.nlm.nih.gov/bioproject) by Gibiome (Seongnam, Republic of Korea). Relative transcript abundance was measured as fragments per kilobase per million mapped reads (FPKM) (37). RNA sequence data were submitted to the NCBI Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo) with accession code GSE200806.

#### Western blotting analysis

After immunization of rabbits with each purified protein in Ab Frontier (Seoul, Republic of Korea), polyclonal antibodies were generated and verified by Western blotting. Western blotting analysis was conducted as described (13).

#### Isothermal titration calorimetry (ITC)

ITC experiments were conducted using a VP-ITC instrument from Malvern Panalytical (Malvern, UK) at 25°C. The wild-type, the T55I mutant, and the 36-bp target DNA were dissolved in 10 mM Tris-HCl buffer (pH 7.4) containing 200 mM KCl, 1 mM tris(2-chloroethyl) phosphate (TCEP) and 0.2 mM EDTA. The solutions were degassed for 3 min before being loaded into the ITC instrument. The concentration of the wild-type and the T55I mutant in the cell was 25  $\mu$ M, while the concentration of the 36-bp target DNA in the syringe was 380 µM. Titration experiments consisting of 25 injections in total were performed with continuous stirring at 307 rpm. The injection volume was 2 µl for the first injection to minimize effects of bubbles and 11 µl for the remaining injections. The initial delay and the reference power were set to 1200 s and 10 µcal/s, respectively. The 36-bp target DNA solution was titrated into the same buffered solution for the heat of dilution. Binding isotherms were displayed after subtracting the dilution heat and baseline correction. Fitting analyses of binding isotherms were carried out using the one-set-of-sites binding model in the MicroCal Origin 7.0 software.

#### Luciferase reporter assay

A luciferase reporter assay to monitor the specific binding of EnfR to the target DNA sequence was designed by using a heterologous bacterial host *E. coli* (38). To construct the EnfR-producing plasmids, wild-type and mutant TON\_1525 (EnfR) genes in the expression plasmids were subcloned into pJK1113, a plasmid containing the arabinose-inducible promoter  $P_{BAD}$  (Supplementary Table S5) (39). The resulting plasmids were designated as the pJH\_EnfR series. To construct synthetic promoters that are repressed by EnfR, we engineered two constitutively expressing bacterial promoters (J23117 and J23110) (40) to have the EnfR-binding site either at the very downstream of or at partially overlapping position with the -10 box for bacterial RNAP (see 'The target DNA sequence and the implication of the N-terminal loop in DNA recognization' of the Results and Discussion section). DNA fragments containing such engineered sequences (Supplementary Table S6) were chemically synthesized and then cloned into the plasmid pBBR\_lux containing a promoterless luxCDABE operon (Supplementary Table S5) (41). The resulting reporter plasmids were designated as the pJH\_road-blocking (if EnfR binds to the downstream of -10 box) or pJH\_steric\_ hindrance (if EnfR binds to the -10 box thereby interfering RNAP binding). Finally, to produce reporter strains, each pJH\_EnfR plasmid was co-transformed with pJH\_roadblocking or pJH\_steric\_hindrance into E. coli DHa cells (Supplementary Table S5).

For the luciferase reporter assay, overnight cultures of the reporter strains were inoculated into the fresh LB medium supplemented with appropriate antibiotics and L-arabinose for EnfR induction. When the cells were grown to an exponential phase (OD<sub>600</sub> of 0.5), 200  $\mu$ l of the culture was transferred into a well of the Nunc<sup>TM</sup> 96-well white/clear bottom plate (Thermo Fisher Scientific, MA). After further incubation at  $37^{\circ}$ C for 3 h, the OD<sub>600</sub> and cellular luminescence were measured using the Spark<sup>™</sup> microplate reader (Tecan, Switzerland). The relative luminescence unit (RLU) was derived by dividing the cellular luminescence with OD<sub>600</sub>. The relative luminescence level was then calculated by normalizing the RLU of each sample with that of the EnfR-uninduced sample [Relative luminescence level = (RLU of the sample / RLU of the EnfR-uninduced sample)  $\times$  100 (%)].

#### **RESULTS AND DISCUSSION**

#### TON\_1525 is a homo-dimer

Proteins engaged in transcriptional regulations are generally homo-oligomers, and thus, we examined the oligomeric state of TON\_1525 through the sedimentation velocity analytical ultracentrifugation (SV-AUC) (Figure 1). Occupying 99.98% of the total protein peak area, it was analyzed to have a molecular weight of 33,800 Da. Consequently, it can be safely concluded that TON\_1525 exists as a homo-dimer in solution.

# TON\_1525 adopts an 'eighth note' symbol-like fold that is assembled into an X-shaped dimer

We determined the 2.8 Å resolution crystal structure of the wild-type TON\_1525 (Supplementary Table S1). There are two monomers in the asymmetric unit of the wild-type TON\_1525 crystals, which is compatible with our demonstration that TON\_1525 is a dimeric protein in solution. The structures of the two monomers are virtually identical; their structures are superposed with the root mean square (r.m.s.) deviation of 0.490 Å for all C $\alpha$  atoms. Monomeric structure of TON\_1525 resembles an 'eighth note ( $\mathcal{N}$ )' symbol in that it can be divided into three parts: a flag, a straight stem and

a note-head (Figure 2A) (hereafter referred to as 'eighth note' fold regulator, EnfR). The flag part contains two Nterminal  $\alpha 1$  and  $\alpha 2$  helices (residues 1–34), the stem part is a long  $\alpha$ 3 helix (residues 35–62) and the note-head corresponds to the C-terminal  $\alpha+\beta$  domain (residues 63–146) in which a three-stranded antiparallel  $\beta$ -sheet ( $\uparrow\beta1-\downarrow\beta2-\uparrow\beta3$ ) is sandwiched by the stem helix ( $\alpha$ 3) on one face and two Cterminal helices ( $\alpha 4$  and  $\alpha 5$ ) on the other face (Figure 2A). The C-terminal  $\alpha+\beta$  note-head domain has a  $\beta_A-\alpha_A-\alpha_B$ - $\beta_{\rm B}$ - $\beta_{\rm C}$  topology (the alphabet subscripts indicate the order of secondary structural elements represented by Greek letters) (Supplementary Figure S1). This domain is unique in that most proteins having  $\beta - \alpha - \alpha - \beta - \beta$  topology among the proteins belonging to the  $\alpha+\beta$  class defined by Structural Classification of Proteins (SCOP) (42) shows  $\beta_A - \alpha_A - \alpha_B - \beta_C - \beta$  $\beta_{\rm B}$  topology, which displayed a reversed order of the second and third  $\beta$ -strands compared with EnfR.

The dimerization of EnfR is mediated mainly by the flag and stem helices with a contact area of  $\sim 2300$  Å<sup>2</sup>, covering one-third of the molecular surface of a monomer. The long stem helices from two monomers form antiparallel Xshaped cross interactions (Figure 2B). Ala52 and Ile56 from a monomer form a hydrophobic core with the same residues from another monomer at the central intersection where the molecular 2-fold axis is located (Figure 2C). Three hydrophobic residues (Ile58, Trp59, and Ile62) at one end of the stem helix in a monomer make extensive contacts with a hydrophobic patch lined by Leu7, Ile12, Leu15, Arg16 (a three-carbon aliphatic chain in its side chain), Leu46, and Ile49 at the peripheral of the core (Figure 2B). To verify the importance of the hydrophobic interactions for the dimeric conformation, an EnfR mutant harboring three point mutations (A52W, I58G and I62G) was designed. The introduction of Trp into position 52 at the hydrophobic core was for steric clashes, and the two Ile  $\rightarrow$  Gly mutations were to reduce contacts with the hydrophobic patch. As expected, SV-AUC revealed that 72.3% of the total protein peak area was estimated to have a monomeric size (Supplementary Figure S2). In addition to hydrophobic dimeric contacts, electrostatic interactions contribute in part to the dimerization. The acidic EEFDE segment (residues 138–142) of the note-head  $\alpha + \beta$  domain in one monomer makes interactions with a basic side of α1 harboring Lys13, Arg16, and Arg20 in the other monomer (Figure 2C).

*Mycobacterium tuberculosis* RslA-bound RNA polymerase sigma factor  $\sigma_4^L$  is the highest structural homolog of monomeric EnfR searched by the *Dali* server (Supplementary Table S7) (43). However,  $\sigma_4^L$  displays a low structural similarity only towards the  $\alpha$ 2-turn- $\alpha$ 3 HTH motif of EnfR (*Dali Z*-score of 8.5, r.m.s. deviation of 1.8 Å) (Supplementary Figure S3 and Table S7). In addition, the dimerization mode of EnfR is not observed among all structural homologs with *Dali Z*-score >6. The unique tertiary and quaternary structures of EnfR resembling the eighth-note indicate clearly that the Tfx family is structurally distinct from other transcription factor families.

In addition to DNA-binding domains (DBD), transcription factors generally have extra domains that mediate oligomerization and/or recognize molecular signals such as small molecules or proteins. For example, TrmB from *Thermococcales* and *Halobacterium salinarum* contains an



**Figure 1.** Determination of the oligomeric state of EnfR (TON\_1525). Sedimentation Velocity Ultracentrifugation Analysis (SV-AUC) of EnfR. *Left panel*: raw sediment velocity profiles recorded using absorbance at 280 nm. *Right panel*: the distribution of sedimentation coefficient *c*(s) from the model. The analytical ultracentrifugation profile of EnfR showed that it exists as a dimer in solution.



Figure 2. The overall structure of monomer, dimer and the dimeric interface of EnfR. (A) Monomeric structure of EnfR. (B) Side view of dimeric EnfR. Hydrophobic interactions between monomers are indicated by black dashed circles, and the hydrophobic patch is presented by white surface. (C) Bottom view of dimeric EnfR. Electrostatic interactions between the acidic EEFDE segment at the C-terminal region of one monomer and basic residues from the other monomer are indicated by two black squares. The black circle in the dimeric center represents a hydrophobic core. The molecular 2-fold axis is represented by a black oval at the central intersection. Each monomer is colored light green and blue, respectively. All spheres indicate the N- and C-termini of monomers. All residues participating in dimeric interactions are in sticks. For brevity, close-up views are shown only for one monomer in (B) and (C).

N-terminal DBD and an effector binding domain (EBD). Sugar such as sucrose and maltotriose binds to EBD and induces a conformational change of TrmB to control the transcription of genes related to sugar metabolism (44,45). In EnfR, the C-terminal  $\alpha+\beta$  note-head domain corresponds to the extra domain. As an effort to get an idea of the functional role of the note-head domain, we searched for structural homologs using the *Dali* server (Supplementary Table S8). A part of the *E. coli* 4-amino-4-deoxychorismate lyase (PDB code: 112L) was identified as the highest structural homolog; the *Z*-score was 6.1, and the r.m.s. deviation was 3.2 Å for 73 matching C<sub> $\alpha$ </sub> atoms. Structural investigation revealed that the homologous region in the lyase is implicated in dimerization, which suggests that the note-head domain of EnfR appears not to function as an EBD.

The absence of EBD, which senses molecular signals and induces structural alteration, suggests that an antirepressor/repressor system (Ant/Rep) might regulate the DNA-binding activity of EnfR. Ant is generally known to bind to DNA-recognition sites of Rep to prevent the formation of the Rep-operator complex (46). There are also novel types of Ant that bind not to DNA-binding sites but to other regions to inactivate Rep (47). In these ways, Ant disrupts the Rep-operator complex, allowing the transcriptional machinery to initiate gene expression. Unfortunately, there is no known *Ant* of EnfR. Consequently, it is a future challenge to reveal how the negative regulation of the *codh* gene cluster by EnfR is relieved.

#### The target DNA sequence and the implication of the Nterminal loop in DNA recognization

To elucidate the DNA-recognition mode of EnfR, first, we mapped the binding sequence for EnfR in the codh promoter region by using the DNase I protection assay. In the previous study, it was confirmed that the purified EnfR binds specifically to the 150-bp long DNA probe containing the promoter region of *codh* gene cluster (24). When the same DNA probe was incubated with the purified EnfR protein, a region extended from -44 to +3 (centered at -21 from the translation start site of TON\_1017, a gene expressed polycistronically with the codh gene TON\_1018) (13,48) was protected from DNase I digestions (Figure 3A). Inspection of this sequence revealed a 36-bp pseudo-palindromic inverted repeat (5'cCaTCtTAaAATcTttttgtttAcATTaTAcGAgGt-3'); 18-bp minimal binding region is underlined. Considering the dimeric structure of EnfR, this pseudo-palindromic sequence is highly likely to be a target site for EnfR-binding.

Afterward, we built an in silico model structure of the EnfR/DNA complex to reveal the DNA-recognition mode of EnfR in structural aspects. Among structural homologs searched by the Dali server (Supplementary Table S7), there were several RNAP  $\sigma^{70}$ -like  $\sigma$  factors that are necessary for the initiation of transcription in bacteria. Consistently, according to the Pfam analyses, the  $\alpha$ 2-turn- $\alpha$ 3 HTH motif (residues 7-49) in EnfR corresponds to the region 4 helixturn-helix (r4-HTH) motif of RNAP  $\sigma^{70}$  (Figure 3B). The  $\alpha$ 2-turn- $\alpha$ 3 HTH motif of EnfR is nicely overlapped onto the r4-HTH motif of *Thermus aquaticus*  $\sigma^A$  fragment with the r.m.s. deviation of 0.587 Å for C $\alpha$  atoms of residues 7-49 (Supplementary Figure S4). Since the r4-HTH motif is responsible for DNA recognition (49), it would appear that the  $\alpha$ 2-turn- $\alpha$ 3 HTH motif is the DNA-binding site in EnfR. In the dimeric structure, two a2-turn-a3 HTH motifs are  $\sim$ 35.8 Å apart from each other (using the C $\alpha$  of Glu42 as a reference point). Since the spacing between two successive major grooves is  $\sim$ 36 Å in the B-form DNA, the two HTH motifs seem to be suitably arranged for the recognition of the target DNA sequence. Based on these perspectives, the complex model structure was generated using the HADDDOCK v2.4 program (see the Materials and Methods section).

The resulting complex model is quite reasonable considering electrostatic complementarity and favorable interactions between EnfR and DNA (Figure 3C and D). As shown in Figure 3C, the positively-charged electrostatic potential of the DNA-binding interface in EnfR is adequate to accommodate the negatively-charged sugar-phosphate backbone of DNA. The N-terminal loop (residues 1–6, referred to as the *N*-6 region hereafter) in the two monomers is situated in such a way to make contact with the minor groove (Figure 3C). In addition, the N-terminal part of the stem helix ( $\alpha$ 3) in the HTH motifs fits into the major groove with three arginine residues (Arg35, Arg43 and Arg44) interacting with DNA. Arg43 and Arg44 extend their side chains toward the bases in the major groove, and Arg35 makes contact with the phosphorous backbone of DNA (Figure 3D). It should be noted that the guanidino group of arginine residues is well known to interact with the base and the backbone of DNA in the protein-DNA complexes (50).

To verify the *in silico* complex model, we made four mutant proteins (a mutant with the R35A replacement, a mutant with a R43A replacement, a mutant with a R44A replacement, and a mutant with N-terminal six residues deleted  $(\Delta N^{1-6})$ ). Notably, the three arginine residues in the HTH motif (Arg35, Arg43, and Arg44) and the Nterminal six residues are directly involved in DNA-binding in the complex model. Point mutations introduced into the solvent-exposed arginine residues and the deletion of the N-6 region at the N-terminus cannot affect the dimeric nature of EnfR, which was verified by AUC experiments in which all the mutants were revealed to be dimers (Supplementary Figure S5). The DNA-binding activities of the four mutant proteins (the R35A mutant, the R43A mutant, the R44A mutant, and the  $\Delta N^{1-6}$  mutant) were analyzed by the electrophoretic mobility shift assay (EMSA) with the 36-bp target DNA sequence. In the case of the R35A and R43A mutants, their DNA-binding activities were drastically decreased, and the R44A mutant seemed to lose the DNA-binding activity (Figure 4). These observations, together with the defect of the  $\Delta N^{1-6}$  mutant in DNA-binding (Figure 4), demonstrated that the N-6 region and the HTH motif take part in DNA recognition as suggested by the in silico complex model.

For in vivo validation of the in vitro mutational studies, we developed two reporter systems: (i) the road-blocking system, of which the target DNA sequence is located between a bacterial constitutive promoter (J23117) (40) and a transcriptional start site of the luciferase (lux) operon and (ii) the steric hindrance system, of which the target sequence is partially overlapped with another type of bacterial constitutive promoter (J20110) (Figure 5A and B) (40). The binding of EnfR to the target sequence in vivo would interfere with the proceeding or binding of bacterial RNAP in the road-blocking or steric hindrance system, respectively. Indeed, when EnfR was induced via arabinose (Figure 5A and B), the luminescence levels were significantly reduced in both systems in an arabinose concentrationdependent manner. If the wild-type EnfR was induced with 0.01% (v/v) of arabinose in either the road-blocking or steric hindrance system, a relative luminescence level was about 35% compared to the uninduced condition. When R35A, R43A, R44A or  $\Delta N^{1-6}$  was induced, the relative luminescence levels were in the range of 55-85%, which indicated that all the tested mutants have defects in DNA binding as shown in EMSA and that the three arginine residues and the N-6 region contribute to DNA-binding (Figure 5C and D).

Furthermore, to examine the effect of base substitutions on EnfR-binding, four base pairs with palindromic relationships in the target DNA sequence were changed; A10-T27  $\rightarrow$  C10-G27, A11-T26  $\rightarrow$  C11-G26, T12-A25  $\rightarrow$  G12-C25 and T14-A23  $\rightarrow$  G14-C23 (Supplementary Figure S6A and Table S3). Then, the influence of the base substitution



**Figure 3.** In silico model of the EnfR/DNA complex. (A) DNaseI footprinting analysis of EnfR-binding site in the promoter of the *codh* gene cluster. The protected region by EnfR and the susceptible region to DNaseI are represented by grey and black boxes, respectively. (B) Domain architecture of EnfR annotated by *Pfam.* (C) Surface representation of the EnfR dimer with the electrostatic potentials in the *in silico* complex model. Positive, negative and neutral electrostatic potentials are colored blue, red and white, respectively. The 36-bp target DNA is represented as a light green ribbon and the minimal EnfR-binding region (5'-AATCTTTTGTTTACATT-3') is indicated by pink. The N-terminal loop (residues 1–6, referred to as the *N*-6 region) is indicated by a yellow dotted square. (D) The interfaces between the dimeric EnfR and the 36-bp DNA. *Left circle:* DNA-contacting residues (Arg35, Arg43 and Arg44) are shown in green sticks. *Right square:* the residues in the *N*-6 region are also presented by sticks. Hydrogen bonds between the side chain hydroxyl group of Thr55 in a monomer and the backbone -NH group of Phe6 in another monomer are presented by black dotted lines. For clarity, the DNA-binding regions in the dimeric EnfR are colored in green and blue for each monomer.



Figure 4. DNA-binding activity of EnfR variants. EnfR (0 ~ 200 nM) was incubated with the Cy5-labeled 36-bp DNA (5 nM).  $\Delta N^{1-6}$  indicates a mutant with the *N*-6 region deleted.



**Figure 5.** Verification of the EnfR-mediated repression *in vivo*. (A and B) EnfR-mediated repression of the reporter luciferase gene expression was verified either in the road-blocking system (A) or the steric hindrance system (B). The EnfR-binding site is located at very downstream of or at a partially overlapping position, respectively, with the –10 box for bacterial RNA polymerase. (C and D) Various mutant EnfRs were induced with 0.01% (v/v) of L-arabinose, and their repressor function was examined via luciferase reporter gene assay using road-blocking (C) or steric hindrance system (D). (E) EnfR-mediated repression was significantly dampened if the EnfR-binding sequence in the promoter was mutated (T14G-A23C). WT indicates the wild-type EnfR. For A, B and E, a schematic diagram of each promoter is shown above the graph. The statistical significance was determined by ordinary one-way ANOVA with multiple comparisons (A, B and E) or by multiple t-tests (C and D) (*ns*, not significant; \*\*\*\**P* < 0.001; \*\**P* < 0.01; \**P* < 0.05). EnfRB, EnfR-binding site; EnfRB\_mt, Enf-binding site with mutation.



**Figure 6.** Effect of the T55I mutation on the DNA-binding activity of EnfR. (**A**) The crystal structure of the T55I mutant (purple) was superposed onto the wild-type (green) in the EnfR-DNA complex model. To clearly understand the movement of the *N*-6 region, N-termini are represented by spheres, and all  $\alpha$ -helices are shown in cylinders. (**B**) Electrophoretic mobility shift assay (EMSA) for the wild-type and the T55I mutant. EnfR (0–250 nM) was incubated with the Cy5-labeled 36-bp DNA (5 nM). The fractions of bound 36-bp dsDNA were plotted against EnfR concentrations (0, 5, 10, 20, 30, 40, 50, 62.5, 75, 100, 125, 150, 200 and 250 nM) with a Hill 1 equation. The data from the wild-type and the T55I mutant are colored black and red, respectively.

on EnfR-binding was examined by EMSA. Among the four substitutions, the T14G-A23C substitution was remarkable in the reduction of EnfR-binding (Supplementary Figure S6B), which was confirmed by the luciferase reporter assay. Compared to the uninduced control, a relative luminescence level in the presence of the T14G-A23C substituted promoter sequence remained at  $\sim$ 85% and  $\sim$ 65% after the wild-type EnfR was induced with 0.01% and 0.1% (v/v) of arabinose, respectively (Figure 5E). In contrast, when the original target sequence was used, the relative luminescence level was decreased to about 39% and 22% at 0.01% and 0.1% (v/v) arabinose, respectively (Figure 5B), which indicated that the T14G-A23C substitution reduced the repression activity of EnfR. Collectively, it can be concluded that EnfR loosely binds to the target DNA sequence harboring the T14G-A23C substitution. The road-blocking system was not employed in this base substitution experiment since the T14G-A23C substitution resulted in no expression of *lux* operon, probably due to the formation of a stable secondary structure in the 5' untranslated region of the transcript.

# Structural explanation for the defect of the T55I mutant in DNA-binding

The transcription level of genes in the *codh* cluster was reported to be affected by the T55I mutation in EnfR (24). To examine the influence of the T55I mutation on the DNAbinding activity of EnfR, EMSA was performed. The Cy5labelled target DNA sequence was titrated as the concentration of the wild-type and T55I mutant increased. In both cases, single retarded bands were observed, and their intensities were increased in a concentration-dependent manner (Figure 6B), which indicates only one dimeric EnfR binds to the target sequence. Remarkably, the T55I mutation seemed to cause a defect in the DNA-binding activity. The intensity of the retarded band in the presence of the T55I mutant was lower than that of the wild-type in EMSA (Figure 6). Consistently, when the T55I mutant was induced in the luciferase reporter assay, the luminescence level was increased compared to in the presence of the wild-type (Figure 5C and D). For quantitative measurements of the negative effect of the T55I mutation on DNA-binding activity, we determined the dissociation constant  $(K_{\rm D})$  of the



Figure 7. Calorimetry-based characterization of intermolecular interactions between EnfR and 36-bp. (A and B) ITC thermograms (upper) and binding isotherms (lower) observed upon titration of 36-bp into wild-type (A) and mutant EnfR (B). Closed circles and solid lines represent the normalized values of the enthalpy change ( $\Delta H$ ) and fit curves, respectively.

wild-type and the T55I mutant towards the target DNA sequence by isothermal titration calorimetry (ITC). The  $K_D$  values of the wild-type and the T55I mutant were calculated to be  $61.8 \pm 20.5$  and  $948.0 \pm 130.0$  nM, respectively, clearly indicating that the DNA-binding affinity of the mutant is much lower (over fifteen-fold) than that of the wild-type. The stoichiometry parameters (*N*) were  $0.94 \pm 0.01$  for the wild-type and  $0.80 \pm 0.01$  for the T55I mutant, which confirmed that one dimeric EnfR binds to the 36-bp target DNA sequence (Figure 7).

To get structural insights into the reduced DNA-binding activity of the T55I mutant, we determined its 3.4 Å resolution crystal structure (Supplementary Table S1). The overall structures of the monomeric and dimeric T55I mutant are virtually identical to those of the wide-type. The r.m.s. deviations of monomers and dimers between the mutant and the wild-type are 0.827 and 0.941 Å, respectively, for all C $\alpha$ atoms (Figure 8). Despite this structural resemblance, however, the T55I mutant has a remarkable structural difference in the orientation and conformation of the *N*-6 region that is engaged in DNA recognition. In the wild-type dimer, the side chain hydroxyl group of Thr55 in a monomer is hydrogen-bonded to the backbone –NH group of Phe6 in another monomer, which directs the *N*-6 region to extend out of the main body (Figure 8). Two benzyl side chains of Phe6 in the *N*-6 regions of two monomers form  $\pi$ - $\pi$  stacking interaction to stabilize the observed conformation of the *N*-6 region. In contrast, in the case of the T55I mutant, the *N*-6 region runs in a different direction due to the loss of the hydrogen bond, which is the consequence of the isoleucine replacement at position 55 (Figure 8). In this conformation of the two *N*-6 regions, the two phenylalanine residues cannot form  $\pi$ - $\pi$  stacking interactions.

To elucidate structural determinants for the dramatic effect of the T55I mutation on DNA-binding activity, we made a complex model between the T55I mutant and DNA by substituting the wild-type structure with the T55I mutant structure in the *in silico* EnfR/DNA complex model.



**Figure 8.** Conformational changes induced by the T55I mutation. (A) Superposition of monomeric structures of the wild-type (green) and the T55I mutant (purple). A black dotted circle indicates the conformational changes in the *N*-6 region between the wild-type and the T55I mutant. (B) *top*: superposition of dimeric structures of the wild-type (each monomer in green colors of different depth) and the T55I mutant (each monomer in purple colors of different depth). *bottom*: black dotted lines pointed out by red arrows indicate hydrogen bonds between the side chain hydroxyl group of Thr55 in a monomer and the backbone -NH group of Phe6 in another monomer. N-termini are presented by spheres.

In the resulting complex model, it was obvious that the *N*-6 region of the mutant clashes with DNA (Figure 6A). Consequently, it can be concluded that the altered conformation of the *N*-6 region induced by the T55I mutation disturbs the DNA-binding activity of the mutant, and the conformation of T55I mutant is not adequate in DNA-binding.

# EnfR is a repressor regulating the expression of the *codh* gene cluster

The archaeal core promoter region contains two consensus sequences, A-T rich TATA-box and purine-rich BRE (25). Transcription factor B (TFB) and TATA-binding protein (TBP) bind to BRE and TATA-box, respectively, and recruit RNAP to form pre-initiation complexes (PIC) and start transcription (51). When we analyzed the 36-bp target DNA sequence resulting from the DNase I protection assay, there was a TATA box-like sequence consisting of TTATA in 15–19 bp upstream from the translation start site of the codh gene cluster. In other words, EnfR binds to the TATA box-like region. Notably, transcription repressors in archaea usually bind to sequences downstream of BRE including TATA-box regions to block the formation of PIC (52). Therefore, the sequence analysis for the promoter region of the codh gene cluster, together with the reduced affinity of the T55I mutant toward the target DNA sequence, suggests that EnfR acts as a transcriptional repressor that controls the expression level of the *codh* gene cluster.

# Changes in diverse gene expression in the MC11 strain promote the cell growth and $H_2$ production under carboxy-dotrophic condition

According to our previous reports (24,35), the T55I mutation in EnfR (156T strain) and the deletion of EnfR ( $\Delta$ 1525 strain) led to significant increase of CO oxidation activity and H<sub>2</sub> production rate, which was accompanied by the increased expression of the *codh* gene cluster. In addition, we have also shown that  $\Delta 1525$  led to expression changes in various genes (35). To further explore the inherent role of EnfR, we performed comparative transcriptome analyses between the wild-type and the MC11 strain harboring the T55I mutation. Sequencing reads were mapped to 1914 protein-coding genes. Among them, significantly ( $\geq$ 2fold) up- or down-regulated genes in the MC11 strain were 101 (5.3%) and 77 (4.0%) genes, respectively (Supplementary Table S9 and S10). The expression of the gene encoding EnfR (TON\_1525) also increased 2.9-fold, suggesting the potential for self-regulation. The differentially expressed genes (DEGs) were assigned to multiple groups of the archaeal clusters of orthologous genes (arCOGs) (Figure 9A).

Groups P (inorganic ion transport and metabolism) and C (energy production and conversion) had high proportions for up-regulated DEGs (18.7 and 11.3%). On the other hand, groups C (energy production and conversion), I (lipid transport and metabolism) and V (defense mechanisms) showed high proportions for down-regulated DEGs (16.7, 11.8 and 10%). It is noteworthy that group C showed an



**Figure 9.** Transcriptome and phenotypic analysis of the MC11 strain. (A) Distribution of differentially expressed genes (DEGs) in the archaeal clusters of orthologous genes (arCOGs). The number indicates the proportion of genes that show significant ( $\geq$ 2-fold) increase (filled bars) or decrease (empty bars) in each group. The one letter code for arCOG categories is as follows: P, Inorganic ion transport and metabolism; C, Energy production and conversion; U, Intracellular trafficking, secretion and vesicular transport; S, Function unknown; R, General function prediction only; Q, Secondary metabolite biosynthesis, transport and catabolism; E, Amino acid transport and metabolism; H, Coenzyme transport and metabolism; L, Replication, recombination and repair; T, Signal transduction mechanisms; J, Translation, ribosomal structure and biogenesis; G, Carbohydrate transport and metabolism; O, Posttranslational modification, protein turnover and chaperones; N, Cell motility; K, Transcription; F, Nucleotide transport and metabolism; V, Defense mechanisms; M, cell wall/membrane/envelope biogenesis and I, Lipid transport and metabolism. (B, C) Expression profile of (de)hydrogenases in transcriptome analysis (**B**) and western blotting analysis (C). The fold change represents the ratio of the expression level of each gene identified from the triplicate analysis between the wild-type and the MC11 strain. (**D**) Cell growth and H<sub>2</sub> production of the wild-type (closed circles) and the MC11 strain (open circles) strains with 100% CO. The data for the wild-type strain was adapted from our previous report (35). Error bars indicate the standard deviations of three independent cultivations of the MC11 strain in this study.

amphoteric pattern with a high proportion of both up- and down-regulated DEGs. Since hydrogenases are known to play an important role in energy metabolism by generating electrochemical ion gradient along with H<sub>2</sub> production or in H<sub>2</sub> metabolism by disposing of reducing equivalents in hyperthermophilic archaea such as Pyrococcus and Thermococcus (34,53,54), the expression change of the genes encoding them was further analyzed. Transcriptome analyses showed that many hydrogenase genes and related dehydrogenase genes were up-regulated in the MC11 strain: frhA encoding Frh hydrogenase  $\alpha$  subunit (1.4-fold), *mbh* encoding membrane-bound hydrogenase (2.2-fold), mch encoding membrane-bound carbon monoxide-dependent hydrogenase (2.3-fold), *sulf1* encoding soluble hydrogenase Sulf1 (1.2-fold) and *codh* encoding carbon monoxide dehydrogenase (2.4-fold) (Figure 9B). On the contrary, the expression level of *sulf2* gene encoding soluble hydrogenase Sulf2 was 0.7-fold down-regulated (data not shown), and mfh2 encoding membrane-bound formate-dependent hydrogenase was significantly down-regulated by 21.5-fold in the MC11 strain (Figure 9B). Expression changes at the transcription level were consistent with the data at the protein level obtained by Western blotting (Figure 9C).

Previously, resting cell assay revealed enhanced CO oxidation activity of the MC11 strain (24). Considering the upregulation of *codh* and *mch* gene expression, it was assumed that carboxydotrophic growth of the MC11 strain would be improved. According to bioreactor experiments, significant enhancement of cell growth and H<sub>2</sub> production of the MC11 strain was observed (Figure 9D). The maximum cell density and H<sub>2</sub> production rate were increased 2.5- and 2.8fold, respectively, compared to those of the wild-type strain.

In this study, whole transcriptome analysis revealed the potential of EnfR to control expression of a broad array of genes at the transcription level. Sulfur-responsive regulator, SurR and *Thermococcales* glycolytic regulator, Tgr/TrmBL1 (55,56), are well-known global regulators in *Thermococcales* species to regulate diverse gene expression associated with energy conservation and central metabolic pathway for glycolysis and gluconeogenesis, respectively (56,57). It is not yet clear whether EnfR can have pleiotropic effects on various cellular processes. However, the distribution of large numbers of DEGs for a wide range of arCOGs implicates that EnfR might have a role in regulating the expression of diverse genes and has the potential to influence various metabolic processes. To investigate whether EnfR binds to promoter regions of DEGs, we selected two up-regulated genes TON\_1582 and TON\_0537 that encode a Na<sup>+</sup>/H<sup>+</sup> antiporter and sulfhydrogenase  $\beta$  subunit SulfI, respectively. In addition, one down-regulated gene TON\_1563 encoding formate dehvdrogenase was selected. According to EMSA experiments (Supplementary Figure S7), retarded bands for the promoter DNA sequences of the three genes were observed in the presence of EnfR, and they disappeared when cold DNA was added, which showed the specific association of EnfR with the three promoter regions. To verify the possibility of EnfR to act as a global regulator, further studies including chromatin immunoprecipitation sequencing and investigation of phenotypic changes of the EnfR mutants under various culture conditions are needed.

# DATA AVAILABILITY

The atomic coordinates and structure factors of the final models have been deposited in the Protein Data Bank with ID codes 8HNO (the wild-type), 8HNP (the T55I mutant).

RNA sequence data were submitted to the NCBI Gene Expression Omnibus (GEO) database (http://www.ncbi. nlm.nih.gov/geo) with accession code GSE200806.

# SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank beamline staffs at Pohang Light Source (Republic of Korea) and Photon Factory (Japan) for their assistance during data collection.

# FUNDING

This work was supported by the project titled 'Development of potential antibiotic compounds using polar organism resources' [20200610, KOPRI Grant PM23030 to S.-S.C.]; Korea Institute of Ocean Science and Technology (KIOST) in-house program [EA0122 to H.S.L.]; National Research Foundation of Korea (NRF) [NRF-2022R1A2C3006924 to S.-S.C., NRF-2022R1A5A102641311 to S.-S.C., RS-2023-00218150 to B.S.K., NRF-2022R1A2C1011793 to Y.-H.L.]; National Research Council of Science & Technology (NST) [CCL22061-100 to Y.-H.L.] grant funded by the Korea government (MSIT). Funding for open access charge: National Research Foundation of Korea (NRF) [NRF-2022R1A2C3006924] grant funded by the Korea government (MSIT).

Conflict of interest statement. None declared.

## REFERENCES

1. Sharma, S., Agarwal, S. and Jain, A. (2021) Significance of hydrogen as economic and environmentally friendly fuel. *Energies*, **14**, 7389–7416.

- 2. British Petroleum (2021) *BP statistical review of world energy*. British Petroleum.
- 3. Ang,T.-Z., Salem,M., Kamarol,M., Das,H.S., Nazari,M.A. and Prabaharan,N. (2022) A comprehensive study of renewable energy sources: classifications, challenges and suggestions. *Energy Strategy Rev.*, **43**, 100939.
- IRENA. (2019) Global energy transformation: a roadmap to 2050 (2019 edition). In: *International Renewable Energy Agency*. Abu Dhabi.
- Ishaq,H., Dincer,I. and Crawford,C. (2022) A review on hydrogen production and utilization: challenges and opportunities. *Int. J. Hydrog. Energy*, 47, 26238–26264.
- Sharma, S. and Ghoshal, S.K. (2015) Hydrogen the future transportation fuel: from production to applications. *Renew. Sust. Energ. Rev.*, 43, 1151–1158.
- Ball, M. and Wietschel, M. (2009) The future of hydrogen opportunities and challenges. *Int. J. Hydrog. Energy*, 34, 615–627.
- Pareek, A., Dom, R., Gupta, J., Chandran, J., Adepu, V. and Borse, P.H. (2020) Insights into renewable hydrogen energy: recent advances and prospects. *Mater. Sci. Energy. Technol.*, 3, 319–327.
- Megía,P.J., Vizcaíno,A.J., Calles,J.A. and Carrero,A. (2021) Hydrogen production technologies: from fossil fuels toward renewable sources. A mini review. *ENFUEM*, 35, 16403–16415.
- Zhang, B., Zhang, S.-X., Yao, R., Wu, Y.-H. and Qiu, J.-S. (2021) Progress and prospects of hydrogen production: opportunities and challenges. J. Electron. Sci. Technol., 19, 100080.
- Chandrasekhar, K., Kumar, S., Lee, B.-D. and Kim, S.-H. (2020) Waste based hydrogen production for circular bioeconomy: current status and future directions. *Bioresour. Technol.*, **302**, 122920.
- Sharma,S., Basu,S., Shetti,N.P. and Aminabhavi,T.M. (2020) Waste-to-energy nexus for circular economy and environmental protection: recent trends in hydrogen energy. *Sci. Total Environ.*, **713**, 136633.
- Kim,M.-S., Bae,S.S., Kim,Y.J., Kim,T.W., Lim,J.K., Lee,S.H., Choi,A.R., Jeon,J.H., Lee,J.-H., Lee,H.S. *et al.* (2013) CO-dependent H<sub>2</sub> production by genetically engineered *Thermococcus onnurineus* NA1. *Appl. Environ. Microbiol.*, **79**, 2048–2053.
- Verhaart, M.R.A., Bielen, A.A.M., Oost, J.v.d., Stams, A.J.M. and Kengen, S.W.M. (2010) Hydrogen production by hyperthermophilic and extremely thermophilic bacteria and archaea: mechanisms for reductant disposal. *Environ. Technol.*, 31, 993–1003.
- van Groenestijn, J.W., Hazewinkel, J.H.O., Nienoord, M. and Bussmann, P.J.T. (2002) Energy aspects of biological hydrogen production in high rate bioreactors operated in the thermophilic temperature range. *Int. J. Hydrog. Energy*, 27, 1141–1147.
- Wiegel, J., Ljungdahl, L.G. and Demain, A.L. (1985) The importance of thermophilic bacteria in biotechnology. *Crit. Rev. Biotechnol.*, 3, 39–108.
- Pawar,S.S. and van Niel,E.W.J. (2013) Thermophilic biohydrogen production: how far are we?*Appl. Microbiol. Biotechnol.*, 97, 7999–8009.
- Pfeifer,K., Ergal,İ., Koller,M., Basen,M., Schuster,B. and Rittmann,S.K.M.R. (2021) Archaea Biotechnology. *Biotechnol. Adv.*, 47, 107668.
- Shao, W., Wang, Q., Rupani, P.F., Krishnan, S., Ahmad, F., Rezania, S., Rashid, M.A., Sha, C. and Md Din, M.F. (2020) Biohydrogen production via thermophilic fermentation: a prospective application of *Thermotoga* species. *Energy*, **197**, 117199.
- 20. Bae,S.S., Kim,Y.J., Yang,S.H., Lim,J.K., Jeon,J.H., Lee,H.S., Kang,S.G., Kim,S.-J. and Lee,J.-H. (2006) *Thermococcus onnurineus* sp. nov., a hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent area at the PACMANUS field. *J. Microbiol. Biotechnol.*, **16**, 1826–1831.
- Bae,S.S., Kim,T.W., Lee,H.S., Kwon,K.K., Kim,Y.J., Kim,M.-S., Lee,J.-H. and Kang,S.G. (2012) H<sub>2</sub> production from CO, formate or starch using the hyperthermophilic archaeon, *Thermococcus* onnurineus. Biotechnol. Lett., 34, 75–79.
- Lim,J.K., Bae,S.S., Kim,T.W., Lee,J.H., Lee,H.S. and Kang,S.G. (2012) Thermodynamics of formate-oxidizing metabolism and implications for H<sub>2</sub> production. *Appl. Environ. Microbiol.*, **78**, 7393–7397.
- Amend, J.P. and Shock, E.L. (2001) Energetics of overall metabolic reactions of thermophilic and hyperthermophilic archaea and bacteria. *FEMS Microbiol. Rev.*, 25, 175–243.

- 24. Lee,S.H., Kim,M.S., Lee,J.-H., Kim,T.W., Bae,S.S., Lee,S.-M., Jung,H.C., Yang,T.-J., Choi,A.R., Cho,Y.-J. *et al.* (2016) Adaptive engineering of a hyperthermophilic archaeon on CO and discovering the underlying mechanism by multi-omics analysis. *Sci. Rep.*, **6**, 22896.
- Lemmens,L., Maklad,H.R., Bervoets,I. and Peeters,E. (2019) Transcription regulators in Archaea: homologies and differences with bacterial regulators. J. Mol. Biol., 431, 4132–4146.
- 26. Hochheimer, A., Hedderich, R. and Thauer, R.K. (1999) The DNA binding protein tfx from *Methanobacterium thermoautotrophicum*: structure, DNA binding properties and transcriptional regulation. *Mol. Microbiol.*, **31**, 641–650.
- 27. An, Y.J., Kim, M.-K., Song, J.M., Kang, M.H., Lee, Y.-H. and Cha, S.-S. (2015) Vapor batch crystallization and preliminary X-ray crystallographic analysis of a cold-active endo-β-1,4-glucanase that was produced through the cold temperature protein expression. *Biodesign*, **3**, 138–142.
- 28. Kabsch, W. (2010) XDS. Acta Crystallogr. D, 66, 125-132.
- 29. Otwinowski,Z. and Minor,W. (1997) In: *Methods Enzymol.* Academic Press, Vol. 276, pp. 307–326.
- Liebschner, D., Afonine, P.V., Baker, M.L., Bunkoczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A.J. et al. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr. D: Struct. Biol., 75, 861–877.
- Cowtan,K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D*, 62, 1002–1011.
- Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D: Biol. Crystallogr.*, 66, 486–501.
- 33. Kim,B.S., Hwang,J., Kim,M.H. and Choi,S.H. (2011) Cooperative regulation of the *Vibrio vulnificusnan* gene cluster by NanR protein, cAMP receptor protein, and *N*-acetylmannosamine 6-phosphate. *J. Biol. Chem.*, 286, 40889–40899.
- 34. Lee,H.S., Kang,S.G., Bae,S.S., Lim,J.K., Cho,Y., Kim,Y.J., Jeon,J.H., Cha,S.-S., Kwon,K.K., Kim,H.-T. *et al.* (2008) The complete genome sequence of *Thermococcus onnurineus* NA1 reveals a mixed heterotrophic and carboxydotrophic metabolism. *J. Bacteriol.*, **190**, 7491–7499.
- 35. Lee,S.H., Lee,S.M., Kang,S.G. and Lee,H.S. (2021) Enhanced H<sub>2</sub> production by deletion of the Tfx family DNA-binding protein in the hyperthermophilic archaeon *Thermococcus onnurineus* NA1. *Int. J. Hydrog. Energy*, **46**, 35189–35197.
- 36. Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. and Ragg, T. (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.*, 7, 3.
- 37. Trapnell,C., Williams,B.A., Pertea,G., Mortazavi,A., Kwan,G., van Baren,M.J., Salzberg,S.L., Wold,B.J. and Pachter,L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, 28, 511–515.
- 38. Hwang,J., Kim,B.S., Jang,S.Y., Lim,J.G., You,D.-J., Jung,H.S., Oh,T.-K., Lee,J.-O., Choi,S.H. and Kim,M.H. (2013) Structural insights into the regulation of sialic acid catabolism by the *Vibriovulnificus* transcriptional repressor NanR. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, E2829–E2837.
- Jong,L. and Sang,C., H. (2014) IscR is a global regulator essential for pathogenesis of *vibriovalnificus* and induced by host cells. *Infect. Immun.*, 82, 569–578.
- 40. Florea, M., Hagemann, H., Santosa, G., Abbott, J., Micklem, C.N., Spencer-Milnes, X., de Arroyo Garcia, L., Paschou, D., Lazenbatt, C., Kong, D. *et al.* (2016) Engineering control of bacterial cellulose production using a genetic toolkit and a new cellulose-producing strain. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E3431–E3440.

- 41. Lenz, D.H., Mok, K.C., Lilley, B.N., Kulkarni, R.V., Wingreen, N.S. and Bassler, B.L. (2004) The small RNA chaperone hfq and multiple small rnas control quorum sensing in *vibrioharveyi* and *vibriocholerae*. *Cell*, **118**, 69–82.
- 42. Andreeva,A., Kulesha,E., Gough,J. and Murzin,A.G. (2020) The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Res.*, 48, D376–D382.
- Holm,L. and Rosenström,P. (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res., 38, W545–W549.
- 44. Krug, M., Lee, S.-J., Boos, W., Diederichs, K. and Welte, W. (2013) The three-dimensional structure of TrmB, a transcriptional regulator of dual function in the hyperthermophilic archaeon *Pyrococcus furiosus* in complex with sucrose. *Protein Sci.*, 22, 800–808.
- Kim, M., Park, S. and Lee, S.-J. (2016) Global transcriptional regulator TrmB family members in prokaryotes. J. Microbiol., 54, 639–645.
- 46. León, E., Navarro-Avilés, G., Santiveri, C.M., Flores-Flores, C., Rico, M., González, C., Murillo, F.J., Elías-Arnanz, M., Jiménez, M.A. and Padmanabhan, S. (2010) A bacterial antirepressor with SH3 domain topology mimics operator DNA in sequestering the repressor DNA recognition helix. *Nucleic Acids Res.*, 38, 5226–5241.
- 47. Kim, M., Kim, H.J., Son, S.H., Yoon, H.J., Lim, Y., Lee, J.W., Seok, Y.-J., Jin, K.S., Yu, Y.G., Kim, S.K. *et al.* (2016) Noncanonical DNA-binding mode of repressor and its disassembly by antirepressor. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E2480–E2488.
- 48. Lee, S.H., Kim, M.-S., Jung, H.C., Lee, J., Lee, J.-H., Lee, H.S. and Kang, S.G. (2015) Screening of a novel strong promoter by RNA sequencing and its application to H<sub>2</sub> production in a hyperthermophilic archaeon. *Appl. Microbiol. Biotechnol.*, **99**, 4085–4092.
- Campbell,E.A., Muzzin,O., Chlenov,M., Sun,J.L., Olson,C.A., Weinman,O., Trester-Zedlitz,M.L. and Darst,S.A. (2002) Structure of the bacterial RNA polymerase promoter specificity σ subunit. *Mol. Cell*, 9, 527–539.
- Luscombe, N.M., Laskowski, R.A. and Thornton, J.M. (2001) Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucleic Acids Res.*, 29, 2860–2874.
- Reeve, J.N. (2003) Archaeal chromatin and transcription. *Mol. Microbiol.*, 48, 587–598.
- 52. Peeters, E., Peixeiro, N. and Sezonov, G. (2013) Cis-regulatory logic in archaeal transcription. *Biochem. Soc. Trans.*, **41**, 326–331.
- 53. Lim,J.K., Mayer,F., Kang,S.G. and Müller,V. (2014) Energy conservation by oxidation of formate to carbon dioxide and hydrogen via a sodium ion current in a hyperthermophilic archaeon. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 11497–11502.
- 54. Sapra, R., Bagramyan, K. and Adams, M.W.W. (2003) A simple energy-conserving system: proton reduction coupled to proton translocation. *Proc. Natl. Acad. Sci. U.S.A.*, 100, 7545–7550.
- Lee, S.-J., Surma, M., Seitz, S., Hausner, W., Thomm, M. and Boos, W. (2007) Characterization of the TrmB-like protein, PF0124, a TGM-recognizing global transcriptional regulator of the hyperthermophilic archaeon *Pyrococcus furiosus*. *Mol. Microbiol.*, 65, 305–318.
- 56. Kanai, T., Akerboom, J., Takedomi, S., van de Werken, H.J.G., Blombach, F., van der Oost, J., Murakami, T., Atomi, H. and Imanaka, T. (2007) A global transcriptional regulator in *Thermococcus kodakaraensis* controls the expression levels of both glycolytic and gluconeogenic enzyme-encoding genes. J. Biol. Chem., 282, 33659-33670
- 57. Lipscomb,G.L., Keese,A.M., Cowart,D.M., Schut,G.J., Thomm,M., Adams,M.W. and Scott,R.A. (2009) SurR: a transcriptional activator and repressor controlling hydrogen and elemental sulphur metabolism in *Pyrococcus furiosus. Mol. Microbiol.*, **71**, 332–349.

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.