Published in final edited form as:

J Mol Biol. 2010 July 16; 400(3): 463–476. doi:10.1016/j.jmb.2010.05.014.

TM0486 from the hyperthermophilic anaerobe *Thermotoga maritima* is a thiamin binding protein involved in response of the cell to oxidative conditions

Zorah Dermoun^{1,2}, Amélie Foulon^{1,2}, Mitchell D. Miller³, Daniel J. Harrington³, Ashley M. Deacon³, Corinne Sebban-Kreuzer^{1,2}, Philippe Roche¹, Daniel Lafitte⁴, Olivier Bornet¹, Ian A. Wilson⁵, and Alain Dolla¹

¹IMR- CNRS, IFR88, 31 chemin Joseph Aiguier, 13402 Marseille cedex 20, France

²Université de Provence, 3 place Victor Hugo 13331 Marseille Cedex 3 France

³Joint Center for Structural Genomics, <http://www.jcsg.org>; Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Stanford University, Menlo Park, CA 94025, USA

⁴INSERM UMR 911 CRO2, Faculté de Pharmacie, 27 bd Jean Moulin, 13385 Marseille Cedex 05, France

⁵Joint Center for Structural Genomics, <http://www.jcsg.org>; Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA, USA

Abstract

Using the COG database, a comparative genome analysis from anaerobic and aerobic microorganisms, was performed with the aim of identifying proteins specific to the anaerobic way of life. Thirty-three COGs were identified, five of which corresponded to proteins of unknown function. We focused our study on TM0486, from *Thermotoga maritima*, that belongs to one of these latter COGs of unknown function, namely COG0011. The crystal structure of the protein was determined at 2 Å resolution. The structure adopts a $\beta\alpha\beta\beta\alpha\beta$ ferredoxin-like fold and assembles as a homotetramer. The structure also revealed the presence of a pocket in each monomer that bound an unidentified ligand

NMR and calorimetric experiments revealed that TM0486 specifically bound thiamin with a K_d of 1.58 μ M, but not hydroxymethyl pyrimidine (HMP), that was implicated previously as a potential ligand. We demonstrated that the TM0486 gene belongs to the same multicistronic unit as TM0483, TM0484 and TM0485. Although these three genes have already been assigned to the transport of HMP, with TM0484 being the periplasmic thiamin/HMP binding protein and TM0485 and TM0483 the transmembrane and the ATPase components, respectively, our results led us to conclude that this operon encodes for an ABC transporter dedicated to thiamin, with TM0486 transporting charged thiamin in the cytoplasm. Given that this transcriptional unit was up-regulated when *T. maritima* was exposed to oxidative conditions, we propose that by chelating cytoplasmic thiamin, TM0486 and, by extension, proteins belonging to COG0011 are involved in the response mechanism to stress that could arise during aerobic conditions.

© 2010 Elsevier Ltd. All rights reserved.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Accession number : Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 1VK8.

Keywords

Thermotoga maritima; Unknown function protein; DUF77; Thiamin binding protein; Oxidative stress

INTRODUCTION

Thermotoga maritima is a hyperthermophilic anaerobic bacterium originally isolated from a geothermally heated marine sediment at Vulcano, Italy 1. It is a chemoorganotrophic thiosulfate and sulfur-reducing organism, which uses preferentially carbohydrates or complex organic matter with an optimal growth temperature of 80°C 1. Members of the order *Thermotogales* are widespread and cosmopolitan, inhabiting various hot spots on the planet (e.g., hot springs, deep-sea and shallow hydrothermal vents, deep aquifers and hot oil reservoirs). Geochemical and microbial analyses of hydrothermal fluids and sediments from subaerial and shallow submarine sites on Vulcano Island reveal that *Thermotoga* species are detected in significant abundance at all the studied sites and that, although strictly anaerobic, they are not precluded from partially oxygenated sediments 2. This situation is also most probably true for *Thermotoga* species that have been isolated from continental hot springs (e.g. *T. thermarum*) which are directly connected to oxidative atmosphere 3. All this information suggests that *Thermotoga* species have to deal with temporary oxygen exposure in the hot ecosystems they inhabit. Recently, it has been reported that *Thermotoga neapolitana*, isolated from a marine hot ecosystem, could not only tolerate low dissolved oxygen partial pressure, but also grow microaerobically 4. Recently, we found out that *Thermotoga maritima* was able to grow in the presence of low concentrations of oxygen of up to 0.5% (vol/vol) in the gas phase 5. However, this capability of aerotolerance is not restricted to *Thermotoga* species. From the last decade, more and more ecological studies have reported the occurrence of strict anaerobic micro-organisms in partially oxygenated biotopes, as an example, the sulphate-reducing bacteria (SRB). Abundance and metabolic activity of SRB in oxic zones of numerous biotopes, including marine and fresh water sediments, as well as cyanobacterial mats, are frequently evaluated as higher than those in neighbouring anoxic zones 6–7. The large number of SRB found in these oxic environments indicates that these organisms can also deal with temporary exposure to elevated oxygen concentrations 8. The nature of the cellular system facilitating oxygen tolerance is actually intriguing. Numerous studies have been performed to get a better understanding of the cellular systems facilitating aerotolerance of anaerobes. Analyses from several anaerobic micro-organisms pointed out conserved molecular systems involved in either oxygen scavenging, such as a flavo-diiron protein as a terminal oxygen reductase, or in the elimination of oxygen reactive species, such as superoxide reductase and NADH-dependent peroxidase 8.

It is now clear that all of these anaerobic organisms that are found in partially oxygenated environments have evolved specific defences against oxidative conditions. Comparative genome analysis from various anaerobes can uncover new systems specific to anaerobes. Specifically, one can assume that, among proteins of unknown function conserved in anaerobes, some are either determinant in anaerobiosis or involved in protection against oxidative conditions. Functional and structural studies on these proteins are thus highly informative. In that respect, *T. maritima* can be considered as a model organism because its complete genome has been sequenced and annotated 9. It consists of 1868 predicted protein encoding genes, with about 33% that do not currently have any assigned function. Moreover, the Joint Center of Structural Genomics (JCSG; www.jcsg.org) initiated a comprehensive structural proteomics project on *Thermotoga maritima* to define the complete set of protein folds and functions that define a single organism 10. To date,

structures corresponding to more than 17% of the complete proteome have been already deposited in the Protein Data Bank. Direct structural coverage of 37% of the expressed soluble proteins (321 unique PDB structures), combined with homology and fold recognition models, has enabled coverage of 72% of the proteome (89% of predicted crystallizable, non-orphan proteins), representing one of the highest structural coverages for a single organism.

In this paper, we report on a comparative genome analysis of aerobic and anaerobic organisms including *T. maritima* using the COGs database (Cluster of Groups of Orthologous proteins) that represents a phylogenetic classification of proteins from complete genomes 11. Each COG includes proteins that are thought to be orthologous, i.e. connected through vertical evolutionary descent 12. The COGs database has become a widely used tool for computational genomics with important applications in the functional annotation of newly sequenced genomes 13–18. This analysis allowed us to select a list of COGs that we propose to be specific for the anaerobic lifestyle and represent new systems. We focused on a COG that corresponded to proteins of unknown function, COG0011, and structurally and functionally studied one member of this COG, TM0486 from *Thermotoga maritima*.

RESULTS

Comparative genomes analysis

Using the COG database (<http://www.ncbi.nlm.nih.gov/COG/new/>), we performed a comparative genome analysis from anaerobic and aerobic microorganisms to identify proteins which were specific to the anaerobes. To select COGs present in all the anaerobic organisms and absent in the aerobic ones, occurrence of COGs in the strict anaerobic organisms *Archaeoglobus fulgidis*, *Methanothermobacter thermautotrophicus*, *Clostridium acetobutylicum*, *Thermotoga maritima* and *Desulfovibrio desulfuricans* G20 was compared to those in strict aerobic organisms *Halobacterium* sp. NRC-1, *Sulfolobus solfataricus*, *Pseudomonas aeruginosa* and *Deinococcus radiodurans*. Thirty-three COGs were identified and considered as specific to anaerobes (Table 1). Twenty-eight correlated with proteins of known function. Among them, COG2033 and COG1773 correspond to the desulfoferrodoxin and the flavo-diiron protein families, respectively, that are involved in superoxide scavenging and oxygen reduction in anaerobic organisms 8. The five remaining COGs represented proteins of unknown function (Table 1) that were assumed to play a crucial role in anaerobiosis. We first focused our attention on proteins belonging to COG0011. A more exhaustive occurrence analysis revealed that this COG is found in microorganisms that grow in anaerobic conditions (strict or facultative anaerobes), and in microaerophilic microorganisms, such as *aquifex aeolicus*.

In order to decipher the functional role of proteins belonging to this COG, we further investigated a member of this COG, TM0486 from *Thermotoga maritima*.

Overall structure of TM0486

The crystal structure of TM0486 was determined (Fig. 1) to a nominal resolution of 2.0 Å using the multiwavelength anomalous diffraction (MAD) method, although some data were included to 1.8 Å. Data collection, refinement and model statistics are summarized in Table 2. The final model consists of one protein tetramer (four protomers – residues 2 – 94 in chains A, B, and D, and, 1 – 94 in chain C), four unknown ligands and 353 water molecules in the asymmetric unit. No electron density was observed for the extra 12-residue N-terminal expression and purification tag in all four protomers and residue 1 in chains A, B, and D. Multiple conformations were modeled for residues A62, B59, C18, D20, and waters B460, D416 and D452. The individual protomers are very similar with pairwise rmsd values

ranging between 0.13 – 0.38 Å for 93 C α atoms and 0.65 – 0.80 Å for all protein atoms. The Matthews coefficient 19 (V_M) is 1.95 Å³/Da with a corresponding estimated solvent content of 37 %. The Ramachandran plot produced by MolProbity shows that 99.7 % of the residues are in favored regions with no outliers (data not shown).

Analysis of the crystal packing suggests that the probable quaternary state in the crystal is a homotetramer with 222 symmetry and is consistent with the results obtained from analytical size exclusion chromatography coupled with static light scattering 20. The tetramer buries ~10,120 Å² and comprises a dimer of dimers (Fig. 1). The most extensive protomer:protomer interface arises from the anti-parallel interaction of strands β 4 from two protomers to form an 8-stranded anti-parallel sheet spanning both protomers. This dimerization interaction (AB and CD) buries ~1200 Å²/protomer and is observed in many $\beta\alpha\beta\beta\alpha\beta$ ferredoxin-like fold proteins, e.g the ACT domain 21. Two dimers pack together to form the tetramer. The 8-stranded β -sheets from the two dimers stack together with an exposed solvent channel down the middle (Fig. 1C). Two additional protomer:protomer interfaces are formed upon tetramerization. The larger interface (AC and BD) buries an additional ~850 Å²/protomer and arises from the interaction of the C-terminal helix (H3) with the β -turn at the ligand binding site from the neighboring protomer. A similar C-terminal helix interaction is seen in other structures with the MTH1187-like fold 22 that contain this extra C-terminal helix. The final interface formed on tetramerization buries ~470 Å²/protomer and is formed primarily through interactions between strand β 1 and helix H1 with strand β 1 and helix H1 the neighboring protomer.

Ligand binding and characterization

Additional electron density was observed in all four protomers in a pocket formed above the β -hairpin between strands β 2 and β 3 and lined with residues from strand β 1, helix H1, as well as helix H3 from a neighboring protomer (Figs. 1A, C). Potential hydrogen bond donors / acceptors near the density include backbone carbonyl groups from Ile8, Val10, Gly40 and Pro41, the amide nitrogen of Val10 and the O γ of Thr44 (Fig. 1B). Based on the shape of the electron density, binding site geometry and surrounding hydrogen bond donors and acceptors, it was not possible to unambiguously assign the identity of this bound ligand. Since this density could not be assigned to either an endogenous metabolite that co-purified from *E. coli* expression or to reagents used in the purification and crystallization, it was modeled in the structure as an unknown ligand (UNL).

Thus, to further investigate the possible identity of the bound ligand, comparison of the 3D structure of TM0486 with other similar structures in the PDB was carried out to search for biologically interesting insights that are not detectable by comparing sequences alone. The ferredoxin-like ($\beta\alpha\beta\beta\alpha\beta$) fold of TM0486 is found in numerous (483) other proteins. (<http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi>). This fold is frequently found among the ACT domains of allosterically regulated enzymes (Aspartate kinase; Chorismate mutase, TyrA-prephenate dehydrogenase) 21, 23, 24 which are related to the regulatory domain of the *E. coli* 3-phosphoglycerate dehydrogenase (3PGDH), and RAM (Regulation of Amino-acid Metabolism) domains involved in the regulation of amino-acid metabolism in prokaryotes 25.

Analysis of the crystal structures of one member of the COG0011 family, which has a similar 222 tetrameric arrangement of ferredoxin-like domains, YkoF from *Bacillus subtilis*, revealed that one thiamin molecule was bound in each domain. YkoF is a dimer with each protomer having two ferredoxin-like ($\beta\alpha\beta\beta\alpha\beta$) domains, resulting in four thiamins bound per dimer (two in each protomer). The authors thus proposed YkoF as a novel thiamin-binding protein²⁶. A search for other similar structures using the DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server/) 27–29 allowed us to retrieve eight similar

structures (Fig. S1, supplementary data) that all contained the same β -hairpin next to a cavity that could potentially accommodate a small, thiamin-like compound. We thus assayed thiamin as potential ligand for TM0486.

Thiamin binding analyses using NMR and ITC

To experimentally assess the binding of thiamin, NMR experiments were performed on the recombinant TM0486 protein produced in *E. coli* grown in minimal media without thiamin. Both thiamin and its derivative thiamin pyrophosphate (TPP) were tested for binding. In waterLOGSY experiments 30, we observed positive signals in the presence of either thiamin or TPP and a negative signal with the reference molecule (DMSO). It indicated that thiamin and TPP bound TM0486 in a similar way (Fig. 2A). Because TM0483, TM0484 and TM0485 were proposed to be related to the *E. coli* K-12 and *B. subtilis* taurine uptake ABC transporter 31, binding of taurine was also assayed. Negative signals observed on WaterLOGSY spectra with taurine or hydroxymethyl pyrimidine (HMP) indicated that TM0486 was unable to bind both ligands (Fig. 2A). The ^1H - ^{15}N HSQC spectrum of TM0486 (Fig. 2B) was well resolved, characteristic of a well-folded protein of low molecular weight that then likely represents a monomer rather than a tetramer. Addition of thiamin or TPP to the protein sample resulted in two effects on the NMR spectrum. The first one was a shift of a large number of peaks in slow exchange on the NMR time scale which was in agreement with a conformational change of the protein upon thiamin or TPP binding. Secondly, significant line broadening was observed that could be assigned to oligomerization of the protein in the presence of the ligand. In the presence of taurine, no modification of the ^1H - ^{15}N HSQC spectrum of the protein was observed, confirming its lack of binding to the protein.

Isothermal titration calorimetry (ITC) was then used to assess the binding affinity of TM0486 for thiamin (Fig. 3). The experimental data perfectly fit the model based on the assumption that each monomer bound one thiamin molecule ($n = 0.92$). The dissociation constant K_d was calculated to be $\Delta H = -11$ kcal/mol. In addition, no HMP binding was detected in agreement with the NMR experiments described above. Calorimetric thermal denaturation of TM0486 showed that the protein tolerated high temperature ($T_m = 116^\circ\text{C}$ with or without thiamin, Fig. S2, supplementary data).

When TM0486 was produced in *E. coli* grown in either minimal medium supplemented with 1 mM thiamin or in rich LB medium, the protein was purified as a tetramer, as deduced from native gel and DLS experiments (Fig. 4). This oligomerization of the protein can be correlated with the low-resolution NMR spectra of the protein (data not shown). In both growth conditions, the cofactors extracted from the purified protein exhibited ^1H -NMR spectra highly similar to thiamin hydrochlorate (Fig. 5). These data allowed us to identify thiamin as the cofactor of TM0486.

Molecular modelling of Thiamin binding

Molecular modelling was performed in order to predict the possible mode of binding of thiamin to TM0486. Thiamin was modelled into the binding cavity of TM0486 by superimposing the three-dimensional structures of C-terminal DUF77 domain of YkoF (1SBR) and TM0486 using VMD multiseq plugin 32 (Fig. 6A). The local fit between the thiamin compound from YkoF and the unidentified ligand from TM0486 was improved using minor manual adjustments and underlined the role of Thr49 and Thr44 from YkoF and TM0486, respectively. It is noteworthy that the three hydrogen bonds responsible for thiamin specificity in YkoF are conserved in the TM0486 model with the unidentified ligand (Fig. 6B). The identity of the unidentified ligand could not be unequivocally assigned in the electron density map. Crystallographic refinement with the UNL replaced with the thiamin

model does not fully account for the observed density. Specifically, the density level for the pyrimidine ring is much higher than the thiazole group and extra density near the C6A atom cannot be modeled by thiamin alone. Attempts to fit other single pyrimidine and purine analogs also did not fully account for the density suggesting that there could be a mixture of compounds in the crystal. Refinement and averaging protocols to enhance the density quality in this region have failed to resolve the ambiguity. Attempts to co-crystallize TM0486 with thiamin have not yielded crystals suitable for diffraction studies. Nevertheless, the available experimental data combined with modeling support the notion that thiamin binds in the TM0486 pocket in a similar orientation to that seen in YkoF.

Genomic context of *TM0486*

T. maritima genome analysis suggested that *TM0486* gene was part of a gene cluster containing three additional genes, *TM0485*, *TM0484* and *TM0483*. These genes are annotated as an ABC transporter-permease protein - cysTW family, a pyrimidine precursor biosynthesis enzyme and a putative ABC transporter-ATP-binding protein, respectively. A similar gene organisation is conserved in *Thermotoga petrophila*, *Clostridium acetobutylicum*, *Desulfitobactrium hafniensae*, *Desulfotomaculum reducens*, *Roseiflexus castenholzii* and *Roseiflexus* sp.RS1 (data not shown).

We tested whether they belonged to the same transcriptional unit by PCR amplifications of the intergenic regions on cDNA templates. Fig. 7 shows that the intergenic regions were amplified between *TM0486* and *TM0485*, *TM0485* and *TM0484*, *TM0484* and *TM0483*. On the contrary, no amplification was obtained between *TM0486* and *TM0487* or *TM0482* and *TM0483*. These data showed that *TM0486* is included in a multicistronic unit composed of four genes *TM0483*, *TM0484*, *TM0485* and *TM0486*.

Effect of oxidative conditions on the expression of *TM0486*

To study the effect of oxidative conditions on the expression of *TM0486*, cells were cultured in three conditions that differed in their oxidative states. The most reduced conditions (termed anaerobic Na₂S-reduced conditions) were obtained by addition of Na₂S into the medium. Increased oxidative conditions were created by omitting Na₂S in the medium (anaerobic non-reduced conditions) and by adding 0.25% (vol/vol) of oxygen in the gas phase (oxygenated conditions) 8. Amounts of *TM0486* transcripts under these various conditions were compared by quantitative real-time PCR. Expression of *TM0486* gene was about 11 and 12 times higher under anaerobic non-reduced and oxygenated conditions than under anaerobic Na₂S-reduced conditions, respectively. Therefore, it appears that *TM0486* gene expression is higher the more oxidative the environmental conditions are. This result suggests a role for *TM0486* in adaptation or response of *Thermotoga maritima* to oxidative conditions.

DISCUSSION

Comparing genomes of several anaerobic organisms led us to identify 33 COGs specific for the anaerobic way of life. Among them, 28 corresponded to proteins with known function, whereas the remaining five corresponded to proteins of unknown function. *TM0486* from *Thermotoga maritima* belongs to one of the latter category of COGs, COG0011. It is noteworthy that this COG is also found in proteomes of some microaerophiles, such as *Aquifex aeolicus* or some aero-anaerobic microorganisms, such as *Saccharomyces cerevisiae*. However, we have demonstrated here that the gene encoding *TM0486* showed higher expression levels when the culture conditions were oxidizing, suggesting a role in adaptation or response of the organism to the presence of oxygen. We propose that this COG0011 protein family is involved in the mechanism by which anaerobic and facultative

or microaerophilic organisms deal with oxidative conditions. Due to the ever increasing numbers of genomes being completely sequenced, the number of putative conserved proteins of unknown function is rapidly growing and the current challenge is to predict or experimentally deduce the activity of these proteins. Genetic environment, bioinformatics and structural analyses are useful for functional prediction and all these tools have been applied in the search here for a function for *Thermotoga maritima* TM0486.

The report of a transcriptional analysis of biofilm formation in *Thermotoga maritima* illustrated that TM0486, as well as TM0483, TM0484 and TM0485, were overexpressed in biofilm compared to planktonic cells 31. In this study, the authors suggested that these genes could be related to *E. coli* K-12 and *B. subtilis* ABC transporter systems for sulfonates coded by *ssuABC* operon and *tauABC* operon, respectively. *TauABC* encodes a taurine ABC transporter while the *ssuABC* operon encodes an alkane sulfonate transport system 33. TM0483, TM0484, TM0485, and TM0486 have been proposed 31 as putative sulphur compounds transport systems possibly involved in sulphur uptake during biofilm formation in *T. maritima*. The need of sulphur could be explained by the iron-sulphur cluster synthesis and repair in biofilms during an oxidative stress as indicated by the up-regulation of the iron-sulphur chaperones *nifSU* in mature biofilms of *E. coli* K-12 34.

Our data clearly demonstrate that taurine does not bind to TM0486. However, specific binding of thiamin is observed with a K_d of 1.58 μ M. In addition, when TM0486 is produced in rich medium or in minimal medium supplemented with thiamin, this ligand co-purifies with the protein.

Our study therefore allows us to propose that TM0486 is a cytoplasmic thiamin binding protein belonging to the same family as YkoF. The binding pocket of thiamin is very similar to the binding pocket of the *B. subtilis* YkoF, but has a higher affinity. It is to be noted that the structural arrangement of these two proteins is quite different from that of the periplasmic thiamin binding protein TbpA, which exhibits a 1000-fold higher affinity for thiamin than YkoF 35–37. Furthermore, Rodionov 38 has proposed that TM0485, TM0484 and TM0483 are part of a transmembrane complex (ThiXYZ) involved in the transport of HMP, where TM0484 (ThiX) is the periplasmic HMP binding protein and TM0485 (ThiY) and TM0483 (ThiZ) the transmembrane and the ATPase components, respectively. However, we demonstrated that TM0486 belongs to the same multicistronic unit as TM0483, TM0484 and TM0485 and that it binds thiamin, but not HMP. Therefore, we cannot exclude that this operon encodes for an ABC transporter dedicated to thiamine where TM0486 would sequester charged thiamin in the cytoplasm.

When *T. maritima* is exposed to oxidative conditions, this transcriptional unit is up regulated. Involvement of thiamin in this response is intriguing. However, it has been already shown that thiamin triphosphate (ThTP) accumulates in response to amino-acid starvation in *E. coli* 39-40. Recently, the presence of a new thiamin derivative, adenosine thiamin triphosphate (AthTP), was also reported to be produced by *E. coli* in response to carbon starvation 41. It was thus proposed that these compounds could initiate a reaction cascade involved in the adaptation of bacteria to highly stringent conditions. The authors concluded that, in bacteria, ThTP and AthTP act as alarmones in response to cellular stress 42–43. In the same way, we propose that TM0486, and by extension proteins belonging to COG0011, are thiamin-binding proteins that are involved in the cell response mechanism to the stresses that arise during anaerobic life.

MATERIALS AND METHODS

Strains, media and growth conditions

Escherichia coli DH5 α (*fhuA2* Δ (*argF-lacZ*)*U169 phoA glnV44* Φ 80 Δ (*lacZ*)*M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) was used for DNA cloning. Transformant cells were selected on Luria-Bertoni (LB) medium in the presence of kanamycin (100 μ g/ml).

Escherichia coli BL21-Codon Plus (DE3)-RIL ($F^- ompT hsdS(r_B^- m_B^-)$ *dcm*⁺ *Tet*^r *gal* 1 (DE3) *endA* *Hte* [*argU ileY leuW Cam*^r]) used for protein productions was cultured in the presence of both chloramphenicol (50 μ g/ml) and kanamycin (100 μ g/ml).

Thermotoga maritima was cultured as previously described 8. Briefly, cell were grown in sealed vials to prevent gas exchange, at 80°C with shaking (110 rpm, Infors Labotron). To obtain reproducible cultures, a constant ratio of 2.6:1 (vol/vol) between the gas (90% N₂, 10% CO₂) and liquid phases was used. Strictly anaerobic growth conditions (termed anaerobic Na₂S-reduced conditions) were obtained by adding 0.25 mM Na₂S in the growth medium, before inoculation. To check the Effect of oxidative conditions on the expression of *TM0486*, anaerobic non-reduced conditions were obtained by the absence of Na₂S added to the medium. Finally, oxygenated conditions were obtained by the addition of either 0.25% or 0.5% (vol/vol) of dioxygen in the gas phase before inoculation.

Quantitative real-time PCR

To adapt cells, *T. maritima* was grown three times in each condition described above before starting the cultures for qRT-PCR experiments. Cells were collected for each condition in the middle of the exponential growth phase by centrifugation (7000g for 20 min. at 4°C) and subsequently washed twice in ice-cold 0.1 M Tris-HCl /0.15 M NaCl (pH 7.6) (10 ml). Cell pellets were subsequently used for RNA extraction. All buffers were made oxygen-free by vigorous bubbling with argon for 20 min. before use. RNA preparation and cDNA synthesis were performed as previously described 8. cDNAs were prepared from three independent cultures for each growth condition.

Amplification and quantification of the cDNAs were carried out with the LightCycler System using the LightCycler FastStart DNA master^{plus} SYBR greenI kit (Roche) according to the manufacturer's instructions. Primers QP486U and QP486D (Table S1, supplementary data) were designed to specifically amplify a portion of *TM0486* gene. The apparatus was programmed for an initial step at 95°C for 480 s, followed by 45 thermal cycles at 95°C for 15 s, 55°C for 15 s, and 72°C for 20 s. The specificities of accumulated products were verified by melting-curve analysis. The Relative Expression Software Tool (REST) was used to calculate the relative expression of each gene in each condition 44, using the 16S RNA gene as reference for normalization. Quantification were performed in triplicate on each cDNA preparation.

RT-PCR for transcriptional analysis

The appropriate primer pair (0.5 μ M each; Table S1, supplementary data) was added to either cDNA or genomic DNA or total RNA, in PCR buffer, together with 100 μ M dNTP and 1U Taq polymerase (Fermentas). The reaction was programmed for 30 cycles at 94°C for 30 s, 53°C for 30 s, 72°C for 40 s and 72°C for 7 min. in a Tgradient thermocycler (Biometra). The PCR products were electrophoresed on 2% agarose gel.

Plasmid construction

Genomic DNA of *Thermotoga maritima* strain MSB8 was prepared from 10 ml of culture using the Wizard genomic Purification kit from Promega. Primers TM0486up (5' CCG TGG AAG ATG GAA GAC TC) and TM0486down (5' CCA CAC AAG CTT GGC TAG CCT)

were designed to specifically amplify the coding region of *TM0486* and introduce both *SacI* and *HindIII* restriction sites. PCR amplification was performed using the Platinum Taq DNA Polymerase High Fidelity" (Roche) in a GeneAmp PCR System 9700 thermal cycler (Applied Biosystems) programmed for an initial step at 96°C for 2 min. followed by a 30 thermal cycles at 96°C for 30 s, 56°C for 30 s and 68°C for 30 s. The amplicon (287 pb) fragment was then digested by *SacI* and *HindIII* and cloned into a pET28b(+) vector digested with the same enzymes. Sequence of the recombinant plasmid pet28-TM0486 was checked by DNA sequencing. *E.coli* BL21 (DE3) RIL was further transformed with pET28-TM0486 for the production of the recombinant protein with an N-terminal His-tag.

TM0486 production and purification

E. coli BL21 (DE3) RIL (pet28-TM0486) was cultured overnight at 37°C, under shaking in minimum medium in the presence of 6 g/l glycerol and 1g/l NH₄Cl supplemented with 0.5 ml of the Widdel oligoelements solution 45, 50 µg/ml chloramphenicol and 100 µg/ml kanamycin. This culture was used to inoculate two 1.5 L flasks containing 500 ml of the same medium. Cells were cultured at 37°C under shaking until OD_{600nm} 1. At that time, protein production was induced by addition of 0.5 mM of IPTG overnight at 37°C.

For NMR experiments, ¹⁵N labeled TM0486 was obtained by growing cells in minimum media, as described above, in the presence of 1 g/l ¹⁵NH₄Cl.

All purification steps were performed at room temperature. Cells were centrifuged for 20 min at 8000g at 20°C, resuspended in 20 ml of binding buffer (25 mM Tris-HCl, 0.3 M NaCl, pH 8) and broken with a French Press (770 PSI). The crude extract was centrifuged for 20 min at 8000 g at 20°C in the presence of 20 µl of 10 mg/ml DNaseI. The supernatant was directly loaded onto a Ni-NTA-Agarose column (Qiagen) at a flow rate of 2 ml/min. His-tagged TM0486 was eluted in binding buffer in the presence of 125 mM imidazole and further dialyzed (membrane Spectropor MW3000 membrane) overnight against 0.3 M NaCl, 25 mM phosphate buffer (pH 6). His-tagged TM0486 was judged pure by PAGE-SDS on 12% polyacrylamide gels. When necessary, the pure protein was concentrated by ultrafiltration on a Millipore Centricon PM 3000.

Cloning, protein expression and crystallization for structure determination

TM0486 (TIGR: TM0486; UniProt: Q9WYV6) was amplified by PCR from *T. maritima* strain MSB8 genomic DNA (ATCC #43589D-5) using *PfuTurbo* (Stratagene) and primer pairs encoding the predicted 5'- and 3'-ends of TM0486. The blunt ends of the PCR product were ligated to vector pMH1 (an arabinose-inducible derivative of Invitrogen's pBAD/Thio vector which encodes an amino-terminal expression and purification tags consisting of the amino acids MGSDKIHHHHHH) predigested with *PmlI* and *PsiI*. The cloning junctions were confirmed by sequencing. Protein expression was performed in a selenomethionine-containing medium using the *Escherichia coli* methionine auxotrophic strain DL41 (F⁻ λ⁻ Met A⁻) 46. Expression was induced by the addition of 0.15% arabinose for 3 h. Lysozyme was added to the culture at the end of fermentation to a final concentration of 1 mg/mL. Bacteria were lysed by sonication after a freeze-thaw procedure in Lysis Buffer [50 mM Tris pH 7.9, 50 mM NaCl, 1 mM MgCl₂, 0.25 mM Tris (2-carboxyethyl) phosphine hydrochloride (TCEP)], and the cell debris was pelleted by centrifugation at 3400g for 60 min. The soluble fraction was applied to a metal chelate affinity resin (Amersham Biosciences) previously charged with nickel and equilibrated with Equilibration Buffer [50 mM potassium phosphate pH 7.8, 0.25 mM TCEP, 10% (v/v) glycerol, 300 mM NaCl] containing 20 mM imidazole. The resin was washed with Wash Buffer [50 mM potassium phosphate pH 7.8, 300 mM NaCl, 40 mM imidazole, 10% (v/v) glycerol, 0.25 mM TCEP], and the protein was eluted with Elution Buffer [20 mM Tris pH 7.9, 300 mM imidazole,

10% (v/v) glycerol, 0.25 mM TCEP]. Buffer exchange was performed to remove imidazole from the eluate, and the protein in Buffer Q [20 mM Tris pH 7.9, 5% (v/v) glycerol, 0.25 mM TCEP] containing 50 mM NaCl was applied to a Resource Q column (Amersham Biosciences) previously equilibrated with the same buffer. Protein was eluted using a linear gradient of 50–500 mM NaCl in Buffer Q. The appropriate fractions were pooled, buffer exchanged in Crystallization Buffer [20 mM Tris pH 7.9, 150 mM NaCl, 0.25 mM TCEP] and concentrated for crystallization assays to 7.5 mg/ml by centrifugal ultrafiltration (Millipore). The protein was crystallized using the nanodroplet vapor diffusion method 47 with standard JCSG crystallization protocols 10. Sitting drops composed of 200 nl protein mixed with 200 nl crystallization solution were equilibrated against a 50 μ l reservoir at 277 K for 33 days prior to harvest. The crystallization reagent consisted of 20% PEG-8000, 0.1 M CHES at pH 9.5, 12% PEG200 (final concentration) was included as a cryoprotectant.

Crystal data collection and refinement

Initial screening for diffraction was carried out using the Stanford Automated Mounting system (SAM) 48 at the Stanford Synchrotron Radiation Lightsource (SSRL, Menlo Park, CA). The crystals were indexed in triclinic space group P1 (Table 2). Multiple-wavelength anomalous diffraction (MAD) data were collected at the Advanced Light Source (ALS, Berkeley, USA) on beamline 8.2.1 at wavelengths corresponding to low energy remote (λ_1) and inflection point (λ_2) of a selenium MAD experiment (Table 2). The data sets were collected at 100K using a Quantum 210 CCD detector (ADSC). Data were integrated and reduced using MOSFLM 49 and then scaled with the program SCALA from the CCP4 suite. Phasing was performed with SOLVE 50 with reflections from 42 – 2.0 Å and gave a mean figure of merit of 0.31 for 8 Se sites. Automated tracing was performed with RESOLVE 51–52. Since the data in the outer shell (1.8–2.0 Å) is incomplete, we have defined the nominal resolution as 2.0 Å, which is the resolution of a dataset that is 100% complete and has the same number of reflections as observed in the current dataset 53. However, 3252 observed reflections between 2.00–1.80 Å (35.2% complete for this shell) are included in the refinement. Model building and completion were done using O 54 and REFMAC 5.1 55 was used for refinement with one TLS group per chain. Data processing and refinement statistics are summarized in Table 2.

NMR experiments

NMR experiments were recorded on a Bruker Avance-II 500 MHz spectrometer at 47°C. WaterLOGSY experiments 30 were carried out on a sample of 0.1 mM TM0486 in the presence of either 0.5 mM thiamin, TPP, HMP or taurine. 0.5 mM DMSO was used as reference buffer. ^1H - ^{15}N HSQC spectra were recorded on ^{15}N -labelled TM0486 (0.1 mM) with and without 0.5 mM thiamin, TPP or taurine.

ITC experiments

ITC experiments were performed at 25°C using a MicroCal MCS instrument. The protein was produced in minimal medium without thiamin and conserved in 25 mM phosphate (pH 6) after purification. Protein concentration was 25 μ M in the calorimetric cell. Each experiment consisted of 7 μ l injections of 500 mM thiamin prepared in the protein dialysis buffer. The heat of dilution was measured by injecting the ligand into the protein-free buffer solution or by additional injections of ligand after saturation. The value obtained was subtracted from the heat of reaction to obtain effective heat of binding. To obtain thermodynamic parameters, titration curves were fitted to a one-site model using MicroCal Origin software. Each experiment was performed twice.

Heat denaturation measurements were carried out on a MicroCal VP-DSC instrument in 0.51 ml cells at a heating rate of 1 C/min. Protein concentrations varied from 1.0 to 2.5 mg/

ml in 25 mM phosphate (pH 6). Curves were corrected for the instrumental baseline obtained by heating the solvent used for protein solution. The reversibility of denaturation was checked routinely by sample reheating after cooling in the calorimetric cell. MicroCal Origin software was used. The accuracy of the calorimetric denaturation enthalpy was 8%. The errors in the temperatures of individual transition obtained by deconvolution of complex endotherms did not exceed 0.4°C.

Cofactor extraction

E. coli BL21 (DE3) RIL (pet28-TM0486) was cultured overnight at 37°C, under shaking in either minimum medium in the presence of 6 g/l glycerol and 1 g/l NH₄Cl and 1 mM thiamin or LB medium, both supplemented with 50 µg/ml chloramphenicol and 100 µg/ml kanamycin. This culture was used to inoculate two 1.5 L flasks containing 500 ml of the same medium. Cells were cultured at 37°C under shaking until OD_{600nm} ~ 1. At that time, protein production was induced by addition of 0.5 mM of IPTG overnight at 37°C. His-tagged TM0486 was purified as described above except that the protein was eluted with 250 mM imidazole from the Ni-NTA column. To extract the cofactor, the pure protein was precipitated with 10% TCA. After centrifugation for 10 min at 13000 rpm, the supernatant was collected and directly analyzed by ¹H-NMR.

Size distribution of TM0486

Dynamic Light Scattering was used to assess the size of purified TM0486 in solution. Light scattering data were collected from 50 µl of TM0486 (6 mg/ml) filtered through 0.22 µm filters (Millex-GV, Millipore) at 47°C using a miniDAWN™ TREOS (Wyatt) in batch mode.

Blue Native polyacrylamide gel electrophoresis (BN-PAGE)

The oligomerization state of TM0486 was checked by running the protein in 14% BN-PAGE 56. 10 µg of protein were resuspended in 20 µL loading buffer (5% Coomassie Blue G250, 500 mM aminocaproic acid). The molecular sizing ladder was prepared by combining bovine serum albumin (60 kDa), Rubredoxin Oxygen Oxydo-reductase (45 kDa) from *Thermotoga maritima* and ferredoxin (10 kDa) from *Desulfovibrio Desulfuricans* Norway. Migration was performed with buffers composed of 50 mM Bis-Tris (pH 7) and 15 mM Bis-Tris (pH 7), 50 mM Tricine, 0.02% Coomassie blue G250 for the anode and the cathode reservoirs, respectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant Sponsor: National Institute of General Medical Sciences, Protein Structure Initiative; Grant Number(s): P50 GM62411, U54 GM074898. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource (SSRL) and the Advanced Light Source (ALS). The SSRL is a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health (National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences). The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

REFERENCES

- Huber R, Langworthy TA, König H, Thömm M, Woese CR, Sleytr UB, Stetter KO. *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. Arch. Microbiol. 1986; 144:324–333.
- Rush A, Walpersdorf E, deBeer D, Gurrieri S, Amùè JP. Microbial communities near the oxic/anoxic interface in the hydrothermal system of Vulcano Island, Italy. Chem. Geol. 2005; 224:169–182.
- Windberger E, Huber R, Trincone A, Fricke H, Stetter KO. *Thermotoga thermarum* sp. nov. and *Thermotoga neapolitana* occurring in African continental solfataric springs. Arch. Microbiol. 1989; 151:506–512.
- Eriksen NT, Nielsen TM, Iversen N. Hydrogen production in anaerobic and microaerobic *Thermotoga neapolitana*. Biotechnol. Lett. 2008; 30:103–109. [PubMed: 17849086]
- Le Fourn C, Fardeau ML, Ollivier B, Lojou E, Dolla A. The hyperthermophilic anaerobe *Thermotoga maritima* is able to cope with limited amount of oxygen : insights into its defence strategies. Environ. Microbiol. 2008; 10:1877–1887. [PubMed: 18397308]
- Musmann M, Ishii K, Rabus R, Amann R. Diversity and vertical distribution of cultured and uncultured Deltaproteobacteria in an intertidal mud flat of the Wadden Sea. Environ. Microbiol. 2005; 7:405–418. [PubMed: 15683401]
- Teske A, Ramsing NB, Habicht K, Fukui M, Küver J, Jörgensen BB, Cohen Y. Sulfate-reducing bacteria and their activities in cyanobacterial mats of Solar lake (Sinai, Egypt). Appl. Env. Microbiol. 1998; 64:2943–2951. [PubMed: 9687455]
- Dolla A, Fournier M, Dermoun Z. Oxygen defense in sulfate-reducing bacteria. J. Biotechnol. 2006; 126:87–100. [PubMed: 16713001]
- Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature. 1999; 399:323–329. [PubMed: 10360571]
- Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, Spraggon G, Klock HE, McMullan D, Shin T, Vincent J, Robb A, Brinen LS, Miller MD, McPhillips TM, Miller MA, Scheibe D, Canaves JM, Guda C, Jaroszewski L, Selby TL, Elsliger MA, Wooley J, Taylor SS, Hodgson KO, Wilson IA, Schultz PG, Stevens RC. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. Proc. Natl. Acad. Sci. USA. 2002; 99:11664–11669. [PubMed: 12193646]
- Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. Science. 1997; 278:631–637. [PubMed: 9381173]
- Fitch WM. Distinguishing homologous from analogous proteins. Syst. Zool. 1970; 19:99–106. [PubMed: 5449325]
- Slesarev AI, Mezhevaya KV, Makarova KS, Polushin NN, Shcherbinina OV, Shakhova VV, Belova GI, Aravind L, Natale DA, Rogozin IB, Tatusov RL, Wolf YI, Stetter KO, Malykh AG, Koonin EV, Kozyavkin SA. The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. Proc. Natl. Acad. Sci. USA. 2002; 99:4644–4649. [PubMed: 11930014]
- Natale DA, Shankavaram UT, Galperin MY, Wolf YI, Aravind L, Koonin EV. Towards understanding the first genome sequence of a crenarchaeon by genome annotation using clusters of orthologous groups of proteins (COGs). Genome Biology. 2001; 1(5):1–19. research0009.
- Nolling J, Breton G, Omelchenko MV, Makarova KS, Zeng Q, Gibson R, Lee HM, Dubois J, Qiu D, Hitti J, Wolf YI, Tatusov RL, Sabathe F, Doucette-Stamm L, Soucaille P, Daly MJ, Bennett GN, Koonin EV, Smith DR. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. J. Bacteriol. 2001; 183:4823–4838. [PubMed: 11466286]
- McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du F, Hou S, Layman D, Leonard S, Nguyen C, Scott K, Holmes A, Grewal N, Mulvaney E, Ryan E, Sun H, Florea L, Miller W, Stoneking T, Nhan M, Waterston R, Wilson RK.

- Complete genome sequence of *Salmonella enterica* serovar *Typhimurium* LT2. *Nature*. 2001; 413:852–817.
17. Makarova KS, Aravind L, Wolf YI, Tatusov RL, Minton KW, Koonin EV, Daly MJ. Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics. *Microbiol. Mol. Biol. Rev.* 2001; 65:44–79. [PubMed: 11238985]
 18. Jordan IK, Kondrashov FA, Rogozin IB, Tatusov RL, Wolf YI, Koonin EV. Constant relative rate of protein evolution and detection of functional diversification among bacterial, archaeal and eukaryotic proteins. *Genome Biology*. 2001; 2:1–9. Research0053.
 19. Matthews BW. Solvent content of protein crystals. *J. Mol. Biol.* 1968; 32:491–497. [PubMed: 5700707]
 20. McMullan D, Canaves JM, Quijano K, Abdubek P, Nigoghossian E, Haugen J, Klock HE, Vincent J, Hale J, Paulsen J, Lesley S. High-throughput Protein Production for X-ray Crystallography and Use of Size Exclusion Chromatography to Validate or Refute Computational Biological Unit Predictions. *J. Struct. Funct. Genomics*. 2005; 6:135–141. [PubMed: 16211510]
 21. Grant GA. The ACT Domain: A small molecule binding domain and its role as a common regulatory element. *J. Biol. Chem.* 2006; 281:33825–33829. [PubMed: 16987805]
 22. Tao X, Khayat R, Christendat D, Savchenko A, Xu X, Goldsmith-Fischman S, Honig B, Edwards A, Arrowsmith CH, Tong L. Crystal structures of MTH1187 and its yeast ortholog YBL001c. *Proteins*. 2003; 52:478–480. [PubMed: 12866058]
 23. Aravind L, Koonin EV. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J. Mol. Biol.* 1999; 287:1023–1040. [PubMed: 10222208]
 24. Chipman DM, Shaanan B. The ACT domain family. *Curr. Opin. Struct. Biol.* 2001; 11:694–700. [PubMed: 11751050]
 25. Ettema TJG, Brinkman AB, Tani TH, Rafferty JB, van der Oost J. A novel ligand-binding domain involved in regulation of amino acid metabolism in Prokaryotes. *J. Biol. Chem.* 2002; 277:37464–37468. [PubMed: 12138170]
 26. Devedjiev Y, Surendranath Y, Derewenda U, Gabrys A, Cooper DR, Zhang RG, Lezondra L, Joachimiak A, Derewenda ZS. The structure and ligand binding properties of the *B. subtilis* YkoF gene product, a member of a novel family of thiamin/HMP-binding proteins. *J. Mol. Biol.* 2004; 343:395–406. [PubMed: 15451668]
 27. Dietmann S, Park J, Notredame C, Heger A, Lappe M, Holm L. A fully automatic evolutionary classification of protein folds: Dali Domain Dictionary version 3. *Nucleic Acids Res.* 2001; 29:55–57. [PubMed: 11125048]
 28. Holm L, Kaariainen S, Wilton C, Plewczynski D. Chapter 5. Using Dali for structural comparison of proteins. *Curr. Protoc. Bioinformatics*. 2006; (Unit 55)
 29. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v.3. *Bioinformatics*. 2008; 24:2780–2781. [PubMed: 18818215]
 30. Dalvit C, Fogliatto G, Stewart A, Veronesi M, Stockman B. WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *J. Biomol. NMR*. 2001; 21:349–359. [PubMed: 11824754]
 31. Pysz MA, Connors SB, Montero CI, Shockley KR, Johnson MR, Ward DE, Kelly RM. Transcriptional analysis of biofilm formation process in the anaerobic, hyperthermophilic bacterium *Thermotoga maritima*. *Appl. Environ. Microbiol.* 2004; 70:6098–6112. [PubMed: 15466556]
 32. Roberts E, Eargle J, Wright D, Luthey-Schulten Z. Multiseq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics*. 2006; 7:382–392. [PubMed: 16914055]
 33. Eichhorn E, van der Ploeg JR, Leisinger T. Deletion analysis of the *Escherichia coli* taurine and alkanesulfonate transport systems. *J. Bacteriol.* 2000; 182:2687–2695. [PubMed: 10781534]
 34. Beloin C, Valle J, Latour-Lambert P, Faure P, Kzreminski M, Balestrino D, Haagensen JA, Molin S, Prensier G, Arbeille B, Ghigo JM. Global impact of mature biofilm lifestyle on *Escherichia coli* K-12 gene expression. *Mol. Microbiol.* 2004; 51:659–674. [PubMed: 14731270]

35. Hollenbach AD, Dickson KA, Washabaugh MW. Overexpression, purification, and characterization of the periplasmic space thiamin-binding protein of the traffic ATPase in *Escherichia coli*. *Protein Expr. Purif.* 2002; 25:508–518. [PubMed: 12182833]
36. Webb E, Claas K, Downs D. *ThiBPQ* encodes an ABC transporter required for transport of thiamine and thiamine pyrophosphate in *Salmonella typhimurium*. *J. Biol. Chem.* 1998; 273:8946–8950. [PubMed: 9535878]
37. Soriano EV, Rajashankar KR, Hanes JW, Bale S, Begley TP, Ealick SE. Structural similarities between thiamin-binding protein and thiaminase-1 suggest a common ancestor. *Biochem.* 2008; 47:1346–1357. [PubMed: 18177053]
38. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. *J. Biol. Chem.* 2002; 277:48949–48959. [PubMed: 12376536]
39. Lakaye B, Wirtzfeld B, Wins P, Grisar T, Bettendorff L. Thiamine triphosphate, a new signal required for optimal growth of *Escherichia coli* during amino acid starvation. *J. Biol. Chem.* 2004; 279:17142–17147. [PubMed: 14769791]
40. Gigliobianco T, Lakaye B, Makarchikov AF, Wins P, Bettendorff L. Adenylate kinase-independent thiamine triphosphate accumulation under severe energy stress in *Escherichia coli*. *BMC Microbiol.* 2008; 8:16. [PubMed: 18215312]
41. Bettendorff L, Wirtzfeld B, Makarchikov AF, Mazzucchelli G, Frédéric M, Gigliobianco T, Gangolf M, De Pauw E, Angenot L, Wins P. Discovery of a natural thiamine adenine nucleotide. *Nat. Chem. Biol.* 2007; 3:321–3212.
42. Bettendorff L, Wins P. Thiamin diphosphate in biological chemistry: new aspects of thiamin metabolism, especially triphosphate derivatives acting other than as cofactors. *FEBS J.* 2009; 276:2917–2925. [PubMed: 19490098]
43. Frédéric M, Delvaux D, Gigliobianco T, Gangolf M, Dive G, Mazzucchelli G, Elias B, De Pauw E, Angenot L, Wins P, Bettendorff L. Thiaminylated adenine nucleotides. Chemical synthesis, structural characterization and natural occurrence. *FEBS J.* 2009; 276:3256–3268. [PubMed: 19438713]
44. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 2001 29e45.
45. Widdel F, Pfennig N. Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. I. Isolation of new sulfate reducing bacteria enriched with acetate from saline environments. *Archiv. Microbiol.* 1981; 129:395–400.
46. Hendrickson WA, Horton JR, LeMaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *Embo J.* 1990; 9:1665–1672. [PubMed: 2184035]
47. Santarsiero BD, Yegian DT, Lee CC, Spraggon G, Gu J, Scheibe D, Uber DC, Cornell RA, Nordmeyer RA, Kolbe WF, Jin J, Jones AL, Jaklevic JM, Schultz PG, Stevens C. An approach to rapid protein crystallization using nanodroplets. *J. Appl. Cryst.* 2002; 35:278–281.
48. Cohen AE, Ellis PJ, Miller MD, Deacon AM, Phizackerley RP. *J. Appl. Cryst.* 2002; 35:720–726.
49. Leslie AGW. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB. Newsletter on Protein Crystallography.* 1992;26.
50. Terwilliger TC, Berendzen J. Automated structure solution for MIR and MAD. *Acta Crystallogr.* 1999; D55:849–861.
51. Terwilliger TC. Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr.* 2003; D59:38–44.
52. Terwilliger TC. Automated side-chain model building and sequence assignment by template-matching. *Acta Crystallogr.* 2003; D59:45–49.
53. Kleywegt GJ. Validation of protein crystal structures. *Acta Crystallogr.* 2000; D56:249–265.
54. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Cryst.* 1991; A47:110–119.
55. Winn MD, Murshudov GN, Papiz MZ. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* 2003; 374:300–321. [PubMed: 14696379]

56. Schägger H, von Jagow G. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal Biochem.* 1991; 199:223–231. [PubMed: 1812789]

Figure 1.

Crystal structure of TM0486. **(A)** Stereo ribbon diagram of one TM0486 protomer color coded from the N-terminus (blue) to the C-terminus (red). Helices (H1–H3) and β -strands (β 1– β 4) are indicated. The location of the unknown ligand (UNL) is indicated with red spheres. **(B)** Diagram showing the secondary-structure elements of TM0486 superimposed on its primary sequence. The labeling of secondary-structure elements is in accord with PDBsum (<http://www.ebi.ac.uk/pdbsum>). For TM0486, the α -helices (H1–H3), β -strands in β -sheet (A), β -turns (β), γ -turns (γ) and β -hairpins are indicated. Residues interacting with the bound unknown ligand are shown with red dots, green triangles indicate residues interacting with the UNL in the same protomer and the blue triangle indicates the residue

interacting with the UNL in the adjacent protomer. (C) Ribbon representation of the TM0486 tetramer showing exhibiting 222 symmetry. Protomers A, B, C and D are depicted in green, cyan, magenta and yellow, respectively and the unknown ligands are shown as red spheres. The β -sheets from protomers A/B and C/D form a continuous anti-parallel β -sheet with strand order $\beta 1$, $\beta 2$, $\beta 3$, $\beta 4$, $\beta 4'$, $\beta 3'$, $\beta 2'$, $\beta 1'$.

Figure 2.

(A) WaterLOGSY experiments performed on TM0486 in the presence of thiamin (A), TPP (B), HMP (C) and taurine (D). The ligand resonances are marked by points and the DMSO resonance is marked by a star.

(B) ^1H - ^{15}N HSQC spectrum of TM0486 in the absence (black) and in the presence of thiamin (red). The protein concentration was 0.1 mM in NaCl 0.3 M, 25mM phosphate buffer (pH 6). Spectra were recorded at 47°C.

Figure 3.

Calorimetric titration of TM0486 with thiamin. Raw data are in the upper plot, and the lower plot represents the integrated heat versus thiamin/protein ratio. The fitted line is calculated assuming one binding site per TM0486 monomer with the following thermodynamic parameters: $n = 0.916$, $K_d = 1.58 \mu\text{M}$, $\Delta H = -11 \text{ kcal/mol}$.

Figure 4.

(A) Native gel of TM0486. TM0486 produced in minimal medium with 1mM thiamin, lane 1, and without thiamin, lane 2. Lane 3 molecular marker made with SAB 60 kDa, ROO from *Thermotoga maritima* (TM0755) 45 kDa, and ferredoxin 10 kDa from *Desulfovibrio desulfuricans* Norway. (B) Size distribution by number of proteins species measured by Dynamic Light Scattering (DLS). Monomeric form of TM0486 produced in minimal medium in the upper plot and TM0486 produced in minimal medium supplemented with thiamine or in rich medium in the lower plot.

Figure 5. ^1H -NMR spectra of thiamin hydrochloride (A) and extracted cofactor from purified TM0486 (B). Spectra were recorded at 25°C.

Figure 6.

Identification of unknown ligand in TM0486 as thiamin. **(A)** Stereo view of the unknown ligand (UNL) bound in the active site of TM0486. The UNL atoms are shown as magenta spheres. The solvent flattened experimental electron density from resolve is shown contoured at 0.7σ in orange. The thiamin from YkoF (1SBR) (after superposition with TM0486) is shown in grey. Chain A is shown, but the density is similar in all four molecules. Residues from helix H1 are omitted for clarity. **(B)** Representation highlighting three conserved hydrogen bonds between thiamin and the polypeptide in TM0486 (left) and YkoF (right).

Figure 7.

(A) RT-PCR analysis of the *TM0486* gene cluster. RNAs were extracted from *T. maritima* grown under anaerobic condition. The cDNA obtained after reverse transcription were further amplified in PCR using the primer pairs described in Table S1 and designated to amplify the intergenic regions between two successive genes. The PCR products were analyzed by electrophoresis on agarose gel. (r) PCRs performed on mRNAs, (g) PCRs performed on genomic DNA, (c) PCRs performed on cDNAs.

Table 1

COGs list specific to the anaerobic way of life. *Thermotoga maritima*, *Archeoglobus fulgidis*, *Methanothermobacter thermautotrophicus*, *Clostridium acetobutylicum* and *Desulfovibrio desulfuricans* G20 were selected as anaerobic microorganisms while *Halobacterium* sp. NRC-1, *Sulfolobus solfataricus*, *Pseudomonas aeruginosa* and *Deinococcus radiodurans* were selected as aerobes. The 33 COGs correspond to the COGs common to the anaerobes and absent in all aerobes. COGs classification was obtained from www.ncbi.nlm.nih.gov/COG/.

COGType	COGNumber	Description
C	COG0426	Uncharacterized flavoproteins
C	COG1141	Ferredoxin
C	COG1151	6Fe-6S prismane cluster-containing protein
C	COG1592	Rubrerhythrin
C	COG1773	Rubredoxin
C	COG1838	Tartrate dehydratase beta subunit/Fumarate hydratase class I, C-terminal domain
C	COG1951	Tartrate dehydratase alpha subunit/Fumarate hydratase class I, N-terminal domain
C	COG2033	Desulfoferrodoxin
D	COG0455	ATPases involved in chromosome partitioning
D	COG1077	Actin-like ATPase involved in cell morphogenesis
E E	COG0253 COG0289	Diaminopimelate epimerase Dihydrodipicolinate reductase
ER	COG1387	Histidinol phosphatase and related hydrolases of the PHP family
G HC	COG1082 COG0543	Sugar phosphate isomerases/epimerases 2-polyphenylphenol hydroxylase and related flavodoxin oxidoreductases
HR	COG1060	Thiamine biosynthesis enzyme ThiH and related uncharacterized enzymes
I	COG1924	Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)
O	COG1067	Predicted ATP-dependent protease
O	COG1180	Pyruvate-formate lyase-activating enzyme
P	COG0798	Arsenite efflux pump ACR3 and related permeases
P	COG1528	Ferritin-like protein
R R	COG0701 COG1237	Predicted permeases Metal-dependent hydrolases of the β -lactamase superfamily II
R	COG1268	Uncharacterized conserved protein
R R	COG1418 COG1489	Predicted HD superfamily hydrolase DNA-binding protein, stimulates sugar fermentation
R	COG1606	ATP-utilizing enzymes of the PP-loop superfamily
R	COG1691	NCAIR mutase (PurE)-related proteins
S	COG0011	Uncharacterized conserved protein
S	COG0391	Uncharacterized conserved protein
S	COG0392	Predicted integral membrane protein
S	COG1433	Uncharacterized conserved protein
S	COG1641	Uncharacterized conserved protein

Table 2

Summary of crystal parameters, data collection, and refinement statistics for TM0486 (PDB: 1vk8)

Space group	P1		
Unit cell parameters	a = 47.65 Å, b = 47.65 Å, c = 49.61 Å, α = 73.8°, β = 62.9°, γ = 73.6°		
Data collection	λ ₁ -Se-Met	λ ₂ -Se-Met	
Wavelength (Å)	1.0332	0.9796	
Resolution range (Å)	41.7-1.8	41.7-1.8	
Number of observations	48,084	53,545	
Number of unique reflections	25,897	28,678	
Completeness (%)	76.1 (23.9) ^a	84.3 (39.4) ^a	
Mean I/σ(I)	10.2 (3.7) ^a	10.4 (3.9) ^a	
R _{merge} on I	0.039 (0.110) ^a	0.049 (0.110) ^a	
Highest resolution shell (Å)	1.85-1.80	1.85-1.80	
Model and refinement statistics			
Resolution range (Å)	45.18-1.80	Data set used in refinement	λ ₁
No. of reflections (total)	25,880	Cutoff criteria	F >0
No. of reflections (test)	1,336	R _{cryst}	0.142
Completeness (% total)	76.1	R _{free}	0.190
Stereochemical parameters			
Restraints(RMSD observed)			
Bond angle (°)	1.44		
Bond length (Å)	0.015		
Average isotropic B-value (Å ²)	23.1 ^c		
ESU based on R _{free} (Å)	0.142		
Protein residues / atoms	373 / 3,036		
Water molecules / unknown ligands	353 / 4		

^aHighest resolution shell.

ESU = Estimated overall coordinate error

$$R_{\text{merge}} = \frac{\sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle|}{\sum_{\text{hkl}} \sum_i I_i(\text{hkl})}$$

$$R_{\text{cryst}} = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}, \text{ where } F_{\text{calc}} \text{ and } F_{\text{obs}} \text{ are the calculated and observed structure factor amplitudes, respectively.}$$

R_{free} = as for R_{cryst}, but for 5.2% of the total reflections chosen at random and omitted from refinement.^bTypically, the number of unique reflections used in refinement is slightly lesser than the total number that were integrated and scaled. Reflections are excluded due to systematic absences, negative intensities and rounding errors in the resolution limits and cell parameters.^cThis value represents the total B that includes TLS and residual B components.