

Research Article

# Structure and function of aerotolerant, multiple-turnover THI4 thiazole synthases

Jaya Joshi<sup>1,\*</sup>, Qiang Li<sup>2,\*</sup>, Jorge D. García-García<sup>1</sup>, Bryan J. Leong<sup>1</sup>, You Hu<sup>2</sup>, Steven D. Bruner<sup>2</sup> and Andrew D. Hanson<sup>1</sup>

<sup>1</sup>Department of Horticultural Sciences, University of Florida, Gainesville, FL, U.S.A.; <sup>2</sup>Department of Chemistry, University of Florida, Gainesville, FL, U.S.A.

Correspondence: Andrew D. Hanson (adha@ufl.edu)



Plant and fungal THI4 thiazole synthases produce the thiamin thiazole moiety in aerobic conditions via a single-turnover suicide reaction that uses an active-site Cys residue as sulfur donor. Multiple-turnover (i.e. catalytic) THI4s lacking an active-site Cys (non-Cys THI4s) that use sulfide as sulfur donor have been biochemically characterized — but only from archaeal methanogens that are anaerobic, O<sub>2</sub>-sensitive hyperthermophiles from sulfide-rich habitats. These THI4s prefer iron as cofactor. A survey of prokaryote genomes uncovered non-Cys THI4s in aerobic mesophiles from sulfide-poor habitats, suggesting that multiple-turnover THI4 operation is possible in aerobic, mild, low-sulfide conditions. This was confirmed by testing 23 representative non-Cys THI4s for complementation of an *Escherichia coli* *ΔthiG* thiazole auxotroph in aerobic conditions. Sixteen were clearly active, and more so when intracellular sulfide level was raised by supplying Cys, demonstrating catalytic function in the presence of O<sub>2</sub> at mild temperatures and indicating use of sulfide or a sulfide metabolite as sulfur donor. Comparative genomic evidence linked non-Cys THI4s with proteins from families that bind, transport, or metabolize cobalt or other heavy metals. The crystal structure of the aerotolerant bacterial *Thermovibrio ammonificans* THI4 was determined to probe the molecular basis of aerotolerance. The structure suggested no large deviations compared with the structures of THI4s from O<sub>2</sub>-sensitive methanogens, but is consistent with an alternative catalytic metal. Together with complementation data, use of cobalt rather than iron was supported. We conclude that catalytic THI4s can indeed operate aerobically and that the metal cofactor inserted is a likely natural determinant of aerotolerance.

## Introduction

Biosynthesis of the adenylated carboxythiazole (ADT) precursor of thiamin is chemically complex and energetically expensive [1,2]. Plants, fungi, and some prokaryotes make ADT via the thiazole synthase THI4, a single-turnover suicide enzyme [3–6]. In a reaction requiring iron (yeast) or zinc (*Arabidopsis*), these THI4s form ADT from NAD, glycine, and a sulfur atom stripped from an active-site Cys residue [3,5,7,8]. The sulfur loss converts Cys to dehydroalanine and irreversibly inactivates the enzyme [3,5] (Figure 1). Such THI4s must, therefore, be replaced after just one reaction cycle, and this — plus the high demand for thiazole [9] — makes THI4 one of the shortest-lived proteins in plant leaves [10,11].

Bioenergetic calculations indicate that the cost of THI4 degradation and resynthesis in plants reduces biomass accumulation by 2–4% [2]. Crop biomass gains of this order could, therefore, result from engineered replacement of a suicide THI4 with a catalytic THI4 that, like most enzymes, mediates thousands of turnovers in its lifetime [12,13]. But do catalytic THI4s that can operate in the mild, aerobic conditions typical of plant cells exist in nature? And if so, what characteristics confer this ability?

\*These authors contributed equally to this work.

Received: 2 August 2021

Revised: 18 August 2021

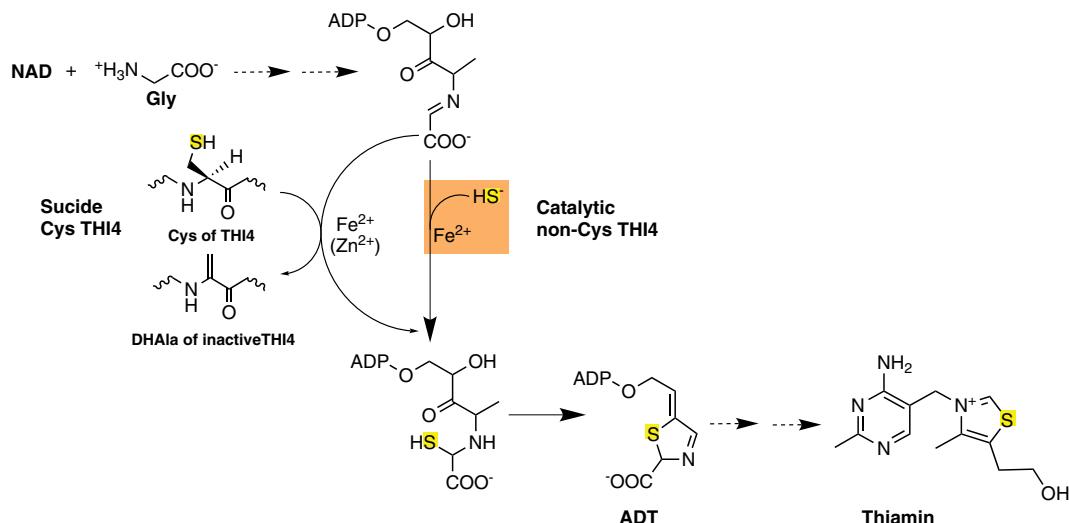
Accepted: 19 August 2021

Accepted Manuscript online:

19 August 2021

Version of Record published:

7 September 2021



**Figure 1. Biosynthesis of the thiazole precursor of thiamin by suicide and catalytic THI4s.**

THI4s form the adenylated carboxythiazole (ADT) precursor of thiamin from NAD, glycine, and a sulfur atom that in yeast and plant THI4s comes from an active-site cysteine residue and in methanococcal THI4s comes from sulfide ( $\text{HS}^-$ ). Sulfur loss from the active-site cysteine leaves a dehydroalanine (DHALa) residue that is not reconverted to cysteine, making yeast and plant THI4s suicide enzymes. In contrast, THI4s that use sulfide mediate multiple turnovers, i.e. are true catalysts.

Catalytic THI4s have been demonstrated biochemically from strictly anaerobic,  $\text{O}_2$ -sensitive, thermophilic methanogens from hydrothermal vents, where sulfide levels are high enough (millimolar) to kill plants and most other organisms [7,14]. These THI4s use sulfide as the sulfur donor, have His in place of Cys in the active site, and prefer iron as cofactor *in vitro* (Figure 1) [7,14]. Indirect evidence suggests that catalytic,  $\text{O}_2$ -sensitive THI4s may also occur in certain plants [5]. An exploratory survey [15] of prokaryote genomes identified THI4s with no active-site Cys (non-Cys THI4s) in several diverse organisms, and preliminary tests showed that two such THI4s complemented an *Escherichia coli* thiazole synthase ( $\Delta\text{thiG}$ ) mutant in aerobic conditions [15]. By indicating that non-CysTHI4s can have at least some activity in mild conditions in the presence of  $\text{O}_2$ , these pilot data prompted further research.

In this work, we deeply surveyed the diversity and genomic contexts of prokaryote non-Cys THI4s and ran complementation assays of thiazole synthase activity on a representative set. In addition, the crystal structure of a THI4 with aerobic complementing activity was determined and this THI4's *in vivo* metal preference was explored. The results implicated the cofactor metal as a determinant of  $\text{O}_2$  tolerance.

## Materials and methods

### Chemicals and media

Chemicals and reagents were from Sigma-Aldrich or Fisher Scientific unless otherwise indicated. MOPS minimal medium was prepared as described [16] except that it was supplemented with the concentrations of micronutrients as specified in [17].

### Bioinformatics

Microbial THI4 sequences were identified in the SEED [18] and UniRef90 [19] databases using *Thermovibrio ammonificans* THI4 as the query sequence. Comparative genomics analyses were performed using SEED and GenBank resources. Sequence similarity networks (SSNs) were constructed by submitting 199 THI4 sequences to the EFI-EST webtool using the FASTA option [20]. An  $E$ -value of  $10^{-5}$  was used to delimit sequence similarity. A final SSN was generated with an alignment score cutoff set such that each connection (edge) represents ~80% sequence identity. In this setting, some sequences remained as singletons. Network layouts were created and visualized using Cytoscape 3.4.

## Knockout strain and clone construction

An *E. coli* MG1655  $\Delta\text{thiG}$  strain was made by recombineering [15,21] using the  $\Delta\text{thiG}$  cassette from the corresponding Keio collection strain [22]. Selected THI4 genes were recoded for expression in *E. coli* or yeast and synthesized by GenScript (Piscataway, NJ) or Twist Biosciences (San Francisco, CA). Recoded nucleotide sequences are given in Supplementary Table S1. For *E. coli*, recoded sequences with an added N-terminal His<sub>6</sub> tag were cloned between the EcoRI and XbaI sites in pBAD24 [23]. For yeast (*Saccharomyces cerevisiae*), the recoded *T. ammonificans* THI4 sequence (preceded by the putative yeast THI4 targeting peptide MSATSTATSTSASQLHLNSTPVTHCLSDGG plus a GG linker) or the native yeast THI4 was inserted into the *CEN6/ARS4* nuclear plasmid carrying the *HIS3* marker and the *TDH3* promoter to drive THI4 expression.

## THI4 protein expression analysis

pBAD24 constructs were introduced into the *E. coli* MG1655  $\Delta\text{thiG}$  strain and single colonies were used to inoculate 3 ml of MOPS medium containing 0.2% (w/v) glycerol, 100 nM thiamin, 100 µg/ml ampicillin, and 50 µg/ml kanamycin. The next day, 25-ml cultures were grown at 37°C in MOPS-glycerol medium with 100 nM thiamin supplementation until OD<sub>600nm</sub> reached 0.8. Cells were then induced by adding 0.02% (w/v) arabinose and incubated for another 4 h at 37°C. Cells were harvested by centrifugation (6000g, 15 min, 4°C), flash-frozen in liquid nitrogen, and stored at -80°C. Cell pellets were extracted by sonicating in 1 ml of 100 mM potassium phosphate (pH 7.2) containing 2 mM β-mercaptoethanol, and separated into soluble and insoluble fractions by centrifugation (17 000g, 10 min, 4°C). Proteins in the pellet were solubilized by boiling for 5 min in 0.5 ml of SDS sample buffer. Aliquots (10 µl) of the insoluble fraction extract or 10-fold-diluted soluble protein extract were separated by SDS-PAGE on 15% gels; proteins were detected by Coomassie Blue staining. The identity of the THI4 bands was confirmed by immunoblotting using anti-His<sub>6</sub> tag antibodies (Thermo Fisher Scientific MA1-21315). The THI4 Coomassie band in soluble and insoluble fractions was quantified using Licor Image Studio Lite software. A 3441-pixel area around the band was used to calculate signal intensity. The method was calibrated using a standard curve for purified recombinant *T. ammonificans* THI4.

## Functional complementation assays in *E. coli* and yeast

For assays in *E. coli*, three independent clones of each construct were used to inoculate 3 ml of MOPS medium containing 0.2% glycerol (w/v), 100 nM thiamin, and 50 µg/ml kanamycin. After incubation at 37°C for 18 h, cells were harvested by centrifugation, washed five times with thiamin-free MOPS medium, resuspended in 500 µl of the same medium, serially diluted in 10-fold steps, and spotted on MOPS medium plates containing 0.2% glycerol, 0.02% arabinose, plus or minus 1 mM Cys. Plates were then incubated at 37°C in aerobic and near-anerobic (N<sub>2</sub> containing ~1 ppm O<sub>2</sub>) conditions as described [15]. Images were captured after 7 d. For complementation assays with *T. ammonificans* THI4 in yeast, three independent clones of strain  $\Delta\text{THI4}$  BY4741 (*MATA his3Δ1 leu2Δ0 met15Δ0 ura3Δ0; THI4Δ::KanMX*) transformed with the *CEN6/ARS4* plasmid alone or containing *T. ammonificans* THI4 or yeast THI4 were used to inoculate 3 ml of synthetic, complete medium (SC; yeast nitrogen base, USBiological cat. no. Y2036), drop-out mix (USBiological cat. no. D9540) minus histidine supplemented with 20 g/l glucose, 5 g/l ammonium sulfate, and 300 nM thiamin. After 48 h of incubation at 30°C and 220 rpm, cells were pelleted (3000g, 5 min), washed five times with thiamin-free SC minus histidine medium, resuspended in the same medium, and used to inoculate 3 ml of thiamin-free SC minus histidine medium to an OD<sub>600nm</sub> of 0.05. Growth was then monitored at OD<sub>600nm</sub>.

## Purification and anaerobic reconstitution of *T. ammonificans* THI4

*T. ammonificans* THI4 cloned in pBAD24 with an N-terminal His<sub>6</sub> tag was transformed into *E. coli* BL21 (DE3). A starter culture (15 ml LB plus 100 µg/ml ampicillin) was inoculated into 6 l of LB; THI4 expression was initiated at an OD<sub>600</sub> of 0.6 with arabinose (0.02% w/v) and incubation was continued at 22°C for 20 h. Cells were then harvested by centrifugation; pellets were resuspended in 50 ml of lysis buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM 2-mercaptoethanol) and lysed in a microfluidizer cell (14 000 psi, M-110L Pneumatic). The lysate was clarified by centrifugation (18 000g, 30 min, 4°C) and applied to a 1-ml Ni-NTA column (HisPur, ThermoFisher Scientific). After incubation at 4°C for 1 h the resin was washed with 50 ml of lysis buffer and eluted with 6 ml lysis buffer plus 250 mM imidazole. The eluate was dialyzed against 1 l of 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 2 mM 2-mercaptoethanol for 12 h and purified by anion exchange

chromatography (HiTrap Q, GE Healthcare) with a linear gradient of 0–500 mM NaCl over 40 min and size-exclusion chromatography (HiLoad 16/600 Superdex 200, GE Healthcare) in 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 2 mM 2-mercaptoethanol. Purified protein fractions were concentrated to 500 µl (14 mg/ml) and transferred to an anaerobic glovebox. Protein reconstitution and crystallization experiments were performed under pure argon and all solutions used were degassed and purged with argon before use. Purified enzyme (500 µl, 14 mg/ml) was incubated with 10 molar equivalents of ferrous ammonium sulfate at 22°C for 30 min, followed by incubation with 10 molar equivalents each of NAD<sup>+</sup> and glycine for 1 h.

### Crystallization and structure solution of *T. ammonificans* THI4

Initial crystallization conditions were obtained using the hanging-drop method at 22°C, with 2 µl reconstituted enzyme and 2 µl of 100 mM HEPES-NaOH, pH 7.5, 200 mM NaCl and 35% 2-methyl-2,4-pentanediol. Optimization of crystal morphology resulted in cubic shaped crystals in ~2 weeks with 100 mM HEPES-NaOH, pH 7.5, 200 mM NaCl, 35% 2-methyl-2,4-pentanediol, 10 mM praseodymium acetate hydrate as the precipitant.

### Data collection, processing, and structure refinement

Diffraction data were collected on beamline 23-ID-D of LS-CAT at Argonne National Laboratory Advanced Photon Source at a wavelength of 1.033 Å. Data were collected at 100°K and processed using XDS [24] to 2.3 Å resolution in space group I121 (Supplementary Table S2). The structure was solved by molecular replacement using a homology model of *Methanococcus igneus* THI4 (PDB: 4Y4N, ~50% sequence identity). A model of *T. ammonificans* THI4 (four monomers per asymmetric unit) was iteratively built by combining AutoBuild (PHENIX [25]) with manual building in COOT [26]. Structure refinement was performed in PHENIX and REFMAC [25,27]. The parameter file for the bound glycine imine intermediate was generated using phenix.elbow [28]. The co-ordinates and structure factors are available from the Protein Data Bank (PDB ID: 7RK0).

## Results

### Diversity of non-Cys THI4s and selection of representatives

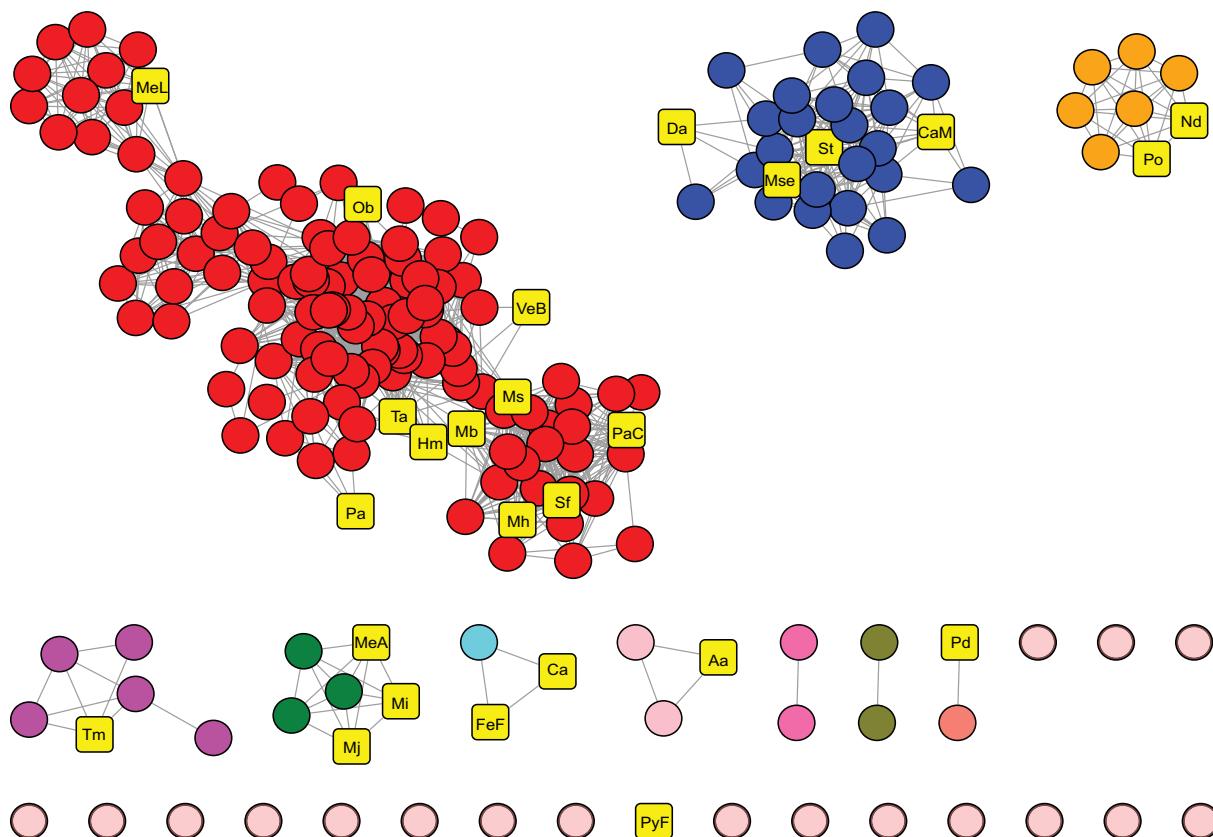
We first surveyed ~1000 prokaryotic THI4 proteins in the SEED [18] and UniRef90 [19] databases using BlastP and non-Cys THI4s [15] as query sequences. After removing entries that were truncated, redundant, or from unidentified organisms, there remained 199 unique sequences in which the active-site Cys position was occupied by His (171 cases) or by Met, Leu, Pro, Ala, Ser, Glu, Asp, Tyr, or Trp (28 cases) (Supplementary Table S3). The 199 sequences shared only 47% identity on average.

To analyze this sequence diversity and help select representatives to test for activity, we built SSNs using the Enzyme Function Initiative webtool [20,29]. The final SSN (*E*-value = 10<sup>-5</sup>, alignment score = 80) contained a series of clusters plus various singletons (Figure 2). Some clusters included non-Cys THI4s that have been tested for activity [7,14,15] but others did not, indicating that much non-Cys THI4 sequence space remained to be sampled.

We selected 26 representative sequences from throughout the SSN and from organisms with different ecophysiology (Figure 2 and Table 1). These sequences were about as diverse (45% average identity) as the whole set of 199 and included THI4s described previously [7,14,15]. Fifteen of the selected sequences were bacterial and 11 were archaeal, eight were from organisms that require or tolerate O<sub>2</sub>, 11 were from mesophiles, 15 were from thermophiles, 22 were from habitats known or likely to be sulfide-rich, and four were from habitats likely to be sulfide-poor (Table 1).

### Genomic context of non-Cys THI4 genes

The selected THI4s all came from genomes that encode ThiE or ThiN (Supplementary Figure S1), indicating capacity to synthesize thiamin from its thiazole and pyrimidine precursors. All genomes also had ThiD and all but one had ThiC (Supplementary Figure S1), respectively, indicating capacity to utilize or produce the pyrimidine precursor. Nearly all the selected THI4s, therefore, came from organisms with complete thiamin synthesis pathways. Only six genomes encoded the alternative thiazole synthase ThiG, so that THI4 was typically the only identifiable endogenous source of the thiazole precursor. Three of the selected archaeal *THI4* genes abut a gene coding for a protein from the TRASH family, whose members bind nonferrous heavy metals, with characterized examples involved in zinc, copper, cadmium, and/or mercury transport/resistance [30–35]



**Figure 2.** Sequence similarity network (SSN) of 199 diverse non-Cys THI4s.

Each node in the SSN corresponds to a single sequence; each edge (gray lines) represents a pair-wise connection between two sequences at a BLAST *E*-value  $< 1 \times 10^{-5}$ . Lengths of edges are not significant, except that tightly clustered groups share more similarity than sequences with only a few connections. The 26 representative sequences selected for testing are shown as yellow squares; organism name abbreviations are as in Table 1.

(Supplementary Figure S1). Similar clustering occurs in many other archaea and bacteria (Figure 3A). The TRASH proteins in these clusters have two adjacent Cys residues in addition to the four canonical metal-binding Cys residues in the TRASH motif (Cys-Xaa<sub>2</sub>-Cys-Xaa<sub>19–22</sub>-Cys-Xaa<sub>3</sub>-Cys) [30] (Figure 3B); these extra Cys could help ligand a six-co-ordinate metal such as cobalt, nickel, or iron. Furthermore, non-Cys THI4 and TRASH genes cluster with genes for proteins from families that transport or metabolize cobalt or nickel [36–39] (Figure 3A). Non-Cys THI4s are thus more strongly genetically associated with nonferrous metals than with iron.

### Functional complementation tests of THI4 activity

The 26 selected THI4s were recoded for *E. coli* and inserted into pBAD24 [23]; the resulting constructs were then introduced into an *E. coli*  $\Delta thiG$  strain [15]. To check THI4 expression, cells were grown in thiamin-supplemented minimal medium and harvested for gel analysis of the soluble and insoluble fractions (Figure 4A,B and Supplementary Figure S2). Of the 26 THI4s, 23 expressed well, with  $\geq 85\%$  in soluble form (Table 1) and were advanced to testing for thiazole synthase activity by complementation.

Complementation tests were run in air or in N<sub>2</sub> containing  $\sim 1$  ppm O<sub>2</sub>, plus or minus supplementation with 1 mM Cys to increase intracellular sulfide level [40]. Sixteen strains showed clear growth in air, particularly when supplemented with Cys while seven did not, and no strain showed clear growth in  $\sim 1$  ppm O<sub>2</sub> but not in air (Figure 4C and Supplementary Figure S3). The enhanced growth with Cys supplementation fits with use of sulfide as sulfur donor [5,15]. The complementation tests in air thus split the THI4s into one group with readily detectable activity (henceforth: aerotolerant THI4s) and another with little or none. Both groups

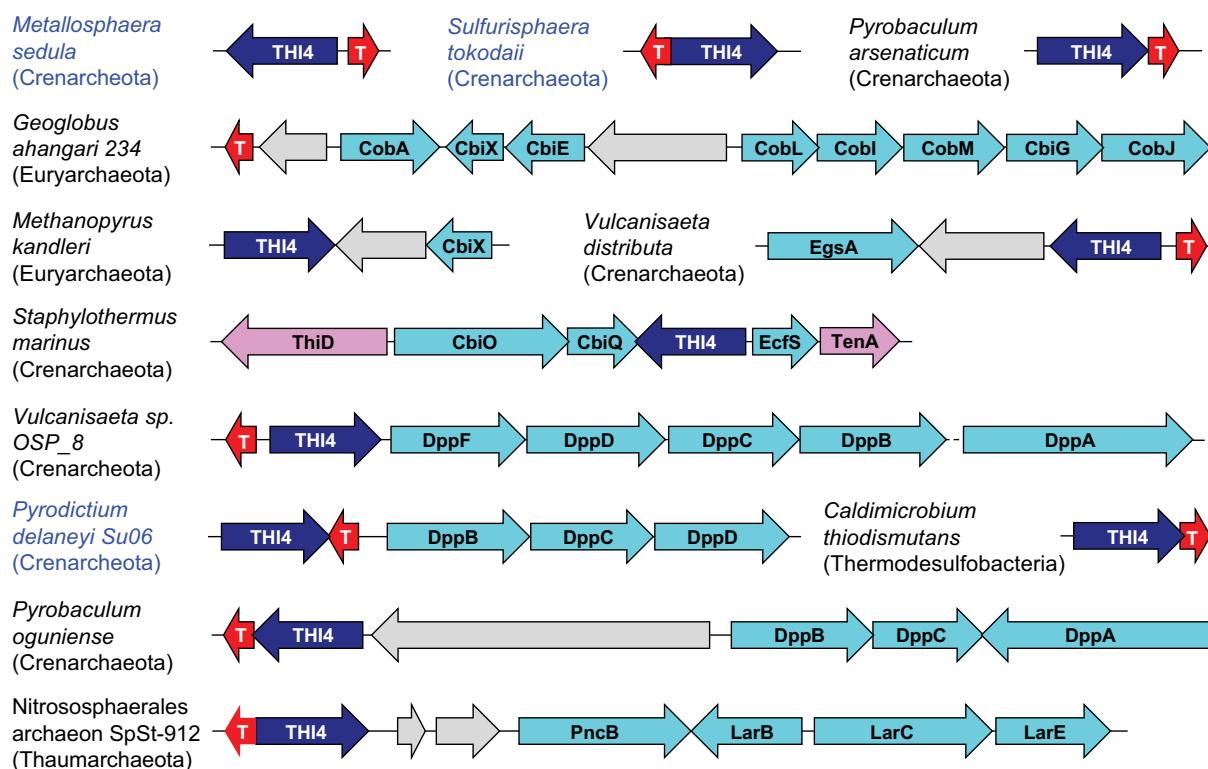
**Table 1.** Ecophysiology of the prokaryote sources of the THI4s selected for testing

| Organism (abbreviation)                      | Natural habitat <sup>1</sup> | O <sub>2</sub> adaptation | Temperature | Sulfide <sup>1,2</sup> |
|--|------------------------------|---------------------------|-------------|------------------------|
| THI4s with readily detectable activity       |                              |                           |             |                        |
| Bacteria                                     |                              |                           |             |                        |
| <i>Fervidicola ferrireducens</i> (FeF)       | Thermal aquifer              | Obligate anaerobe         | Thermophile | Low                    |
| <i>Thermovibrio ammonificans</i> (Ta)        | Hydrothermal vent            | Anaerobe                  | Thermophile | High                   |
| <i>Hippea maritima</i> (Hm)                  | Hydrothermal vent            | Anaerobe                  | Thermophile | High                   |
| <i>Thermotoga maritima</i> (Tm)              | Hot marine sediment          | Anaerobe                  | Thermophile | High                   |
| <i>Caldanaerovirga acetigignens</i> (Ca)     | Hot springs                  | Anaerobe                  | Thermophile | High                   |
| <i>Mucinivorans hirudinis</i> (Mh)           | Leech gut                    | Anaerobe                  | Mesophile   | High                   |
| <i>Parabacteroides chinchillae</i> (PaC)     | Chinchilla gut               | Anaerobe                  | Mesophile   | High                   |
| <i>Verrucomicrobia bacterium</i> (VeB)       | Soil                         | Anaerobe                  | Mesophile   | Low                    |
| <i>Pseudoramibacter alactolyticus</i> (Pa)   | Abscesses                    | Anaerobe                  | Mesophile   | High                   |
| <i>Marinilabilia salmonicolor</i> (Ms)       | Marine mud                   | Facultative aerobe        | Mesophile   | High                   |
| <i>Saccharicrinis fermentans</i> (Sf)        | Marine mud                   | Facultative aerobe        | Mesophile   | High                   |
| <i>Candidatus Marinimicrobia</i> (Mb)        | Sea water column             | Facultative aerobe        | Mesophile   | Low                    |
| Archaea                                      |                              |                           |             |                        |
| <i>Pyrodictium delaneyi</i> (Pd)             | Hydrothermal vent            | Obligate anaerobe         | Thermophile | High                   |
| <i>Methanococcus jannaschii</i> (Mj)         | Hydrothermal vent            | Obligate anaerobe         | Thermophile | High                   |
| <i>Methanococcus aeolicus</i> (MeA)          | Marine sediments             | Obligate anaerobe         | Mesophile   | High                   |
| <i>Methanococcus igneus</i> (Mi)             | Hydrothermal vent            | Anaerobe                  | Thermophile | High                   |
| THI4s with little to no detectable activity  |                              |                           |             |                        |
| Bacteria                                     |                              |                           |             |                        |
| <i>Candidatus Omnitrophica</i> (Ob)          | Hydrothermal vent            | Anaerobe                  | Thermophile | High                   |
| <i>Poribacteria</i> sp. WGA-A3 (Po)          | Sponge symbiont              | Facultative aerobe        | Mesophile   | Low                    |
| <i>Nitrospira defluvii</i> (Nd)              | Sewage sludge                | Aerobe                    | Mesophile   | High                   |
| Archaea                                      |                              |                           |             |                        |
| <i>Desulfurococcus amylolyticus</i> (Da)     | Hot springs                  | Obligate anaerobe         | Thermophile | High                   |
| <i>Candidatus Aenigmarchaeota</i> (Aa)       | Hot springs                  | Anaerobe                  | Thermophile | High                   |
| <i>Pyrolobus fumarii</i> (PyF)               | Hydrothermal vent            | Facultative microaerobe   | Thermophile | High                   |
| <i>Metallosphaera sedula</i> (Mse)           | Acidic hot springs           | Aerobe                    | Thermophile | High                   |
| Poorly expressed archaeal THI4s <sup>3</sup> |                              |                           |             |                        |
| <i>Sulfurisphaera tokodaii</i> (St)          | Hot springs                  | Obligate aerobe           | Thermophile | High                   |
| <i>Methanofollis liminatans</i> (MeL)        | Wastewater bioreactor        | Obligate anaerobe         | Mesophile   | High                   |
| <i>Caldivirga maquilingensis</i> (CaM)       | Acidic hot spring            | Facultative microaerobe   | Thermophile | High                   |

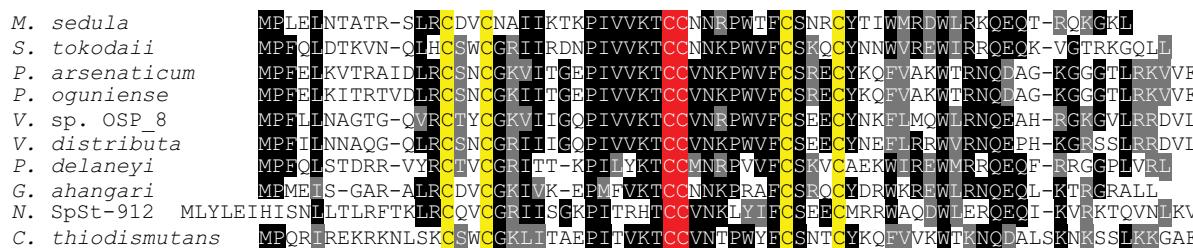
<sup>1</sup>Information was extracted or inferred from literature reporting the isolation or description of the organisms.<sup>2</sup>Broad estimates. Low sulfide: low micromolar range. High sulfide: high micromolar to low millimolar range.<sup>3</sup>Expressed poorly in soluble form in *E. coli* and therefore not analyzed further.

included THI4s from mesophiles and thermophiles, aerobes and anaerobes, and organisms from high- and low-sulfide habitats (Table 1). There was hence no evident correlation between THI4 aerotolerance and source organism ecophysiology. Neither was aerotolerance correlated with the residue (His, Met, Leu, Tyr, or Ser) that replaces Cys in the active site or with the number of Cys and Met residues, which can affect oxidative instability [41,42] (Supplementary Table S3). To summarize: the complementation data establish that some catalytic THI4s operate quite well (although not optimally, see below) in aerobic conditions but do not suggest why others do not. One of many possible causes is that the redox environment in *E. coli* differs enough from that in

**A**



**B**

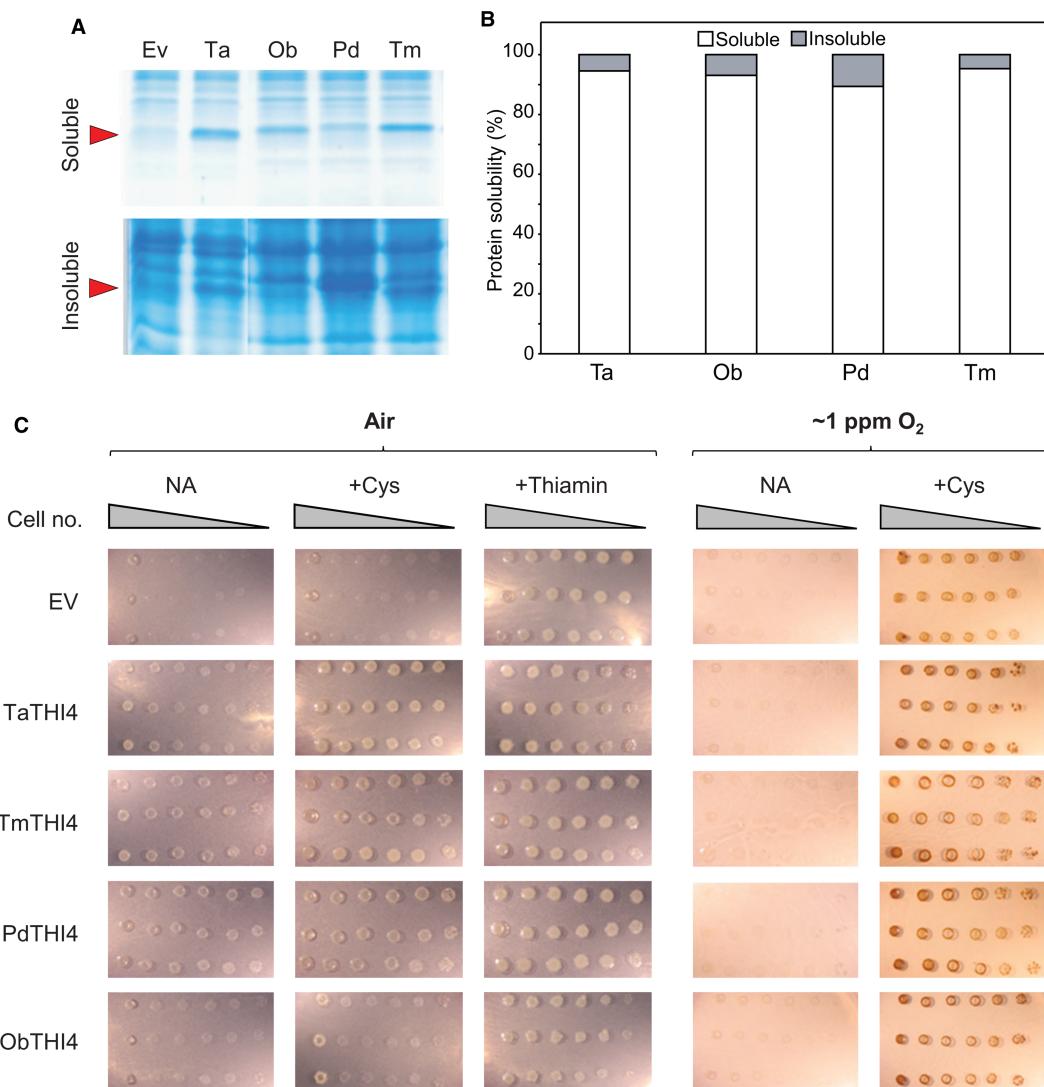


**Figure 3. Association of THI4 genes with TRASH genes and cobalt- or nickel-related genes.**

(A) Chromosomal gene clustering arrangements. Names of organisms whose THI4s were tested are in blue. T, TRASH protein gene. Genes colored aqua encode homologs of nickel- and/or cobalt-related proteins: CobA, uroporphyrin-III C-methyltransferase, involved in cobalamin and sirohaem synthesis; CbiX, sirohydrochlorin nickel/cobalt chelatase, which inserts nickel or cobalt into tetrapyrroles; CbiE, involved in cobalamin synthesis; a cobalamin synthesis operon encoding CobL, CobI, CobM, CbiG, CobJ, plus (not shown) CbiC and CbiD; CbiO, CbiQ, and EcfS, the A + A', T, and S components of a cobalt energy coupling-factor (ECF) transporter. DppA-D,F, subunits of ABC transporters for oligopeptides, nickel, or (rarely) cobalt; LarBCE, nickel-pincer nucleotide (NPN) cofactor synthesis genes; PncB, synthesis enzyme for the NPN precursor NaAD; EgsA, glycerol-1-phosphate dehydrogenase, which has a nickel or zinc cofactor; gray genes have no known nickel or cobalt associations. Genes colored lilac encode thiamin synthesis or salvage enzymes. (B) Alignment of the TRASH proteins encoded by genes in part A. The Cys residues of the extended TRASH motif (Cys-Xaa<sub>2</sub>-Cys-Xaa<sub>19-22</sub>-Cys-Xaa<sub>3</sub>-Cys) are in yellow; the two extra Cys residues are in red.

the THI4 source, especially for extremophiles (Table 1) [43–45], to disrupt protein disulfide formation or metal insertion and coordination [45–47] and thus prevent THI4 expression in its native form.

It is important to note that supplementation with thiamin markedly stimulated the growth of all strains (compare NA and + Thi columns in air in Figure 4C and Supplementary Figure S3). This stimulation shows that the *in vivo* activity of THI4s with complementing activity did not fully meet the demand for ADT despite



**Figure 4. Soluble expression and functional complementation tests of non-Cys THI4s.**

(A) Gel analysis of soluble and insoluble expression in *E. coli* of representative THI4s. Soluble and insoluble fractions of cells were run on 15% gels, stained with Coomassie blue, and scanned to quantify the THI4 band, for which purified *Thermovibrio ammonificans* (Ta) THI4 served as a marker (arrow). Organism abbreviations are as in Table 1. (B) Quantification of the solubility (%) of the THI4s from (A). (C) Tests of functional complementation of an *E. coli*  $\Delta\text{thi}G$  strain by the THI4s from (A) or the empty vector (EV). TaTHI4, TmTHI4, and PdTHI4 had clearly detectable activity in air while ObTHI4 did not; none had more than slight activity in  $\sim 1 \text{ ppm O}_2$ . Overnight cultures of three independent clones per construct were 10-fold serially diluted and spotted on plates of MOPS minimal medium containing 0.2% (w/v) glycerol and 0.02% (w/v) arabinose with no additions (NA) or plus 1 mM Cys or 100 nM thiamin. The medium used for culture in  $\sim 1 \text{ ppm O}_2$  contained 40 mM nitrate as electron acceptor. Cells were cultured in air or under  $\text{N}_2$  containing  $\sim 1 \text{ ppm O}_2$ . Images were captured after incubation at 37°C for 7 d. The high background in the  $\sim 1 \text{ ppm O}_2$  +Cys treatment is due to staining of the inoculum cells.

their high expression levels (Figure 4A and Supplementary Figure S2), i.e. that there is room for activity improvement, at least when no Cys is supplied. We revisit this point later.

### Protein structure of the aerotolerant *T. ammonificans* THI4

As structural data for THI4s with little or no complementing activity in air would be hard to interpret in terms of aerotolerance alone since failure has other possible causes, we explored molecular features associated with

aerotolerance by determining the crystal structure of the complementation-active THI4 from *T. ammonificans* (TaTHI4) (PDB: 7RK0). This bacterium is an anaerobe but is likely to be intermittently exposed to O<sub>2</sub> in its habitat near the oxic/anoxic interface in hydrothermal vents, and encodes enzymes to detoxify reactive oxygen species (ROS), including catalase/peroxidase, cytochrome *c* peroxidase, and cytochrome *bd* complex [48,49]. This makes the TaTHI4 structure a potentially informative ‘routinely O<sub>2</sub>-exposed’ comparison with the structures of ‘never O<sub>2</sub>-exposed’ THI4s from the archaeal methanogens *M. igneus*, *M. jannaschii*, and *Methanothermococcus thermolithotrophicus* [7,50]. Like *T. ammonificans*, these organisms are anaerobic thermophiles from hydrothermal vents [48,51] but, unlike *T. ammonificans*, they inhabit the anoxic region of the vent plume and lack the ROS defense genes present in *T. ammonificans* as well as heme- and manganese-catalase genes (Supplementary Table S4). Furthermore, *M. igneus* and *M. jannaschii* THI4s had similar complementing activity in *E. coli* to TaTHI4 (Supplementary Figure S3), meaning that differences among their recombinant protein structures are unlikely to be artifacts of misfolding in the heterologous host. In comparing the TaTHI4 structure with those of the three archaeal methanogens we made the reasonable assumption that these non-Cys THI4s are all catalytic, as *M. jannaschii* THI4 has been shown to be [14].

Protein crystals of TaTHI4 diffracted to 2.3 Å resolution in space group I121 and the structure was solved by molecular replacement (Supplementary Table S2) using *M. igneus* THI4 as a search model (PDB: 4Y4N) [7]. In common with the methanogen THI4s, TaTHI4 is an overall homooctamer with four monomers per asymmetric unit (Figure 5A). The octamer is tightly packed as a two-layer ring structure with approximate dimensions 73 × 86 Å (height × width) enclosing a ~26 Å diameter pore. The monomer structure (Figure 5B) consists of a long α-helix (residues 6–24) and a barrel-like core domain sandwiched by helix bundles (Figure 5C). The barrel-like core comprises a central five-stranded β-sheet with β6, β2, β1, β12, β13 topology, flanked by a twisted antiparallel β-sheet (β7, β10, β11) connecting α6 and a four-helix bundle (α9, α12, α2, α5). The barrel core is capped by a variation of the Rossmann fold β-hairpin motif. The C-terminal topology is similar to the canonical Rossmann fold with an inversion of strands. TaTHI4 is thus structurally homologous to the previously solved methanogen THI4s [7,50].

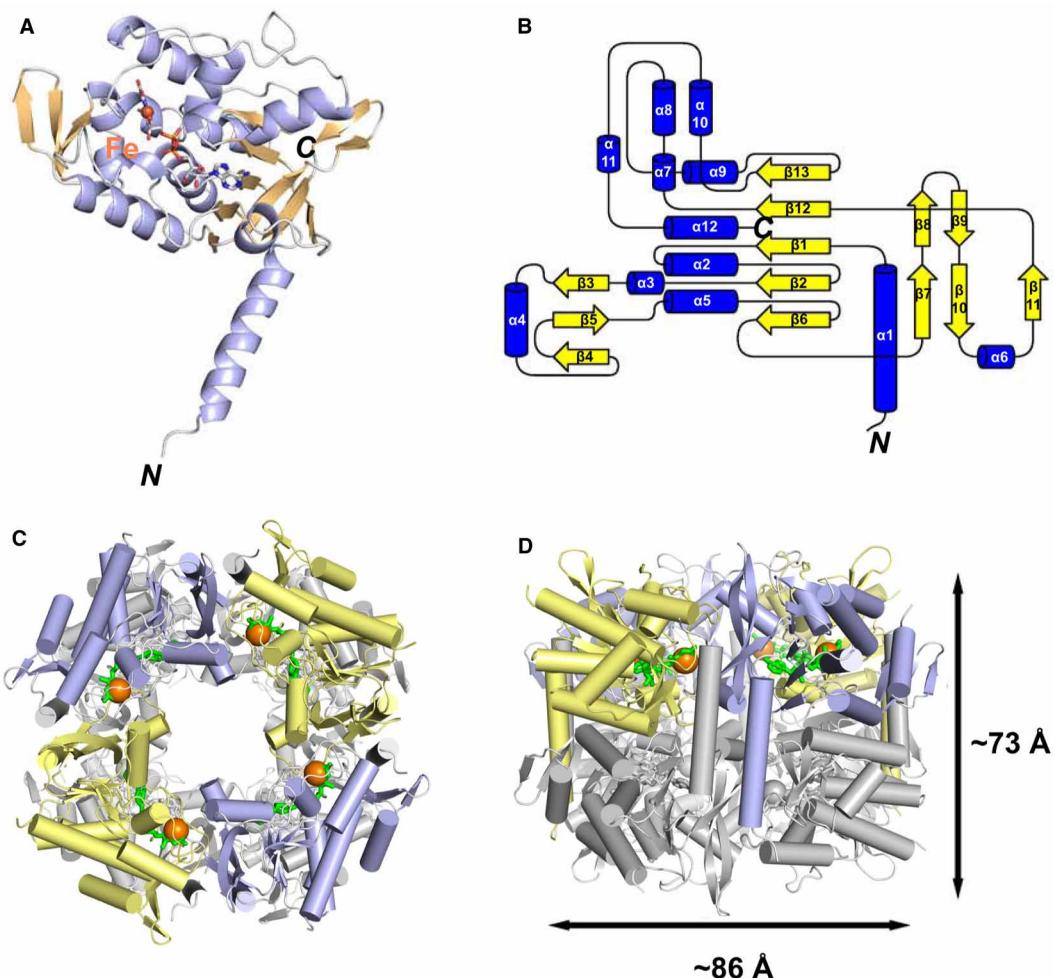
The enzyme active site lies in a channel formed by two helix bundles and is capped by a loop from an adjoining monomer. Clear density for the TaTHI4 treated with NAD and glycine shows the expected glycine imine intermediate [7] in the active site (Figure 6A). This intermediate adopts an elongated conformation with the adenine moiety positioned in the barrel-like core domain. Binding interactions between the intermediate and TaTHI4 include a hydrogen bond between N6 of adenine and the side chain of Ser178, a bidentate hydrogen bond with the 2' and 3' hydroxyls and Glu56, a second hydrogen bond between Arg58 and the 2' hydroxyl, along with a stabilizing interaction of the glycine imine carboxylate and Arg239 (Figure 6A). Based on our reconstitution conditions, the bound metal in the active site can be assigned as ferrous iron, Fe (II). The iron is co-ordinated in a pseudo-octahedral geometry with three sites occupied by the pincer-type ligand of the bound glycine imine, axial ligands of His175 and Asp160, and a predicted water molecule (Figure 6B,C).

To probe the role of the Met158 residue that in TaTHI4 occupies the position of the sacrificial Cys, we changed Met158 to His, Lys, or Cys and assayed complementing activity (Supplementary Figure S4). The mutant enzymes performed similarly to wild-type TaTHI4 in both air and ~1 ppm O<sub>2</sub> atmospheres, with or without Cys supplementation. This result confirms the inference from natural variation in this residue (Supplementary Table S3) that it has no role in catalysis. That the Cys form did not have greater activity in the absence of Cys supplementation suggests that it cannot operate efficiently in suicide mode and still depends on sulfide as sulfur source, as was the case for *M. jannaschii* THI4 [7].

### Investigation of TaTHI4 metal preference *in vivo* in *E. coli* and yeast

As we could not obtain quantifiable *in vitro* activity [14] from reconstituted TaTHI4, we sought to test metal cofactor preference *in vivo*. Adding 100 μM cobalt, nickel, zinc, or manganese to the medium of *E. coli* expressing TaTHI4 (or five other THI4s) did not detectably affect complementing activity in air (Supplementary Figure S5). This result does not indicate which metal is preferred, but does show that the supply of this metal is not limiting in *E. coli* grown with normal trace metal supplementation.

We also tested metal preference using yeast, which resembles *E. coli* in having native cobalt-dependent enzymes and cobalt uptake systems [52] but differs in having no native nickel enzymes or high-affinity nickel uptake system and in needing an added nickel transporter to express a foreign nickel enzyme in active form [53]. We, therefore, tested TaTHI4 for the ability to complement a yeast ΔTHI4 strain. The observed



**Figure 5. Structure of the aerotolerant non-Cys THI4 from *Thermovibrio ammonificans*.**

(A) Ribbon diagram of the TaTHI4 monomer with bound glycine imine shown as sticks;  $\alpha$ -helices are colored blue,  $\beta$ -strands are colored beige; Fe(II) is shown as a brown sphere. (B) Topology diagram of TaTHI4 with  $\alpha$ -helices in blue and  $\beta$ -strands in yellow. (C,D) The biologically relevant TaTHI4 homooctamer shown in ribbon representation in alternate top and side views with the metal Fe(II) indicated as orange spheres and the glycine imine as green sticks. One tetramer is blue/yellow and the other is colored gray.

complementation (Supplementary Figure S6) proves that TaThi4 has access to its metal cofactor and implies that this metal is more likely to be cobalt than nickel if it is not iron.

## Discussion

The biochemically characterized catalytic THI4s come from archaea of anoxic, sulfide-rich, reducing environments, use an O<sub>2</sub>-sensitive sulfur donor (sulfide) and cofactor (ferrous iron), and are thermophilic [7,14]. It was consequently *a priori* doubtful whether catalytic prokaryotic THI4s could function in mild, fully aerobic conditions, and recent evidence [5] that putatively catalytic plant THI4s are expressed only at severely hypoxic stages of seed development reinforced this doubt.

However, our survey of THI4 sequences and their provenances began to dispel the doubt because non-Cys THI4s confidently predicted from genomic context to be functional thiazole synthases were found in aerobic or aerotolerant organisms from mild environments. Confirmation that certain non-Cys THI4s — the majority of those tested, in fact — can function in air came from complementation experiments, which also indicated that these THI4s use sulfide, or a sulfide metabolite, as sulfur donor.

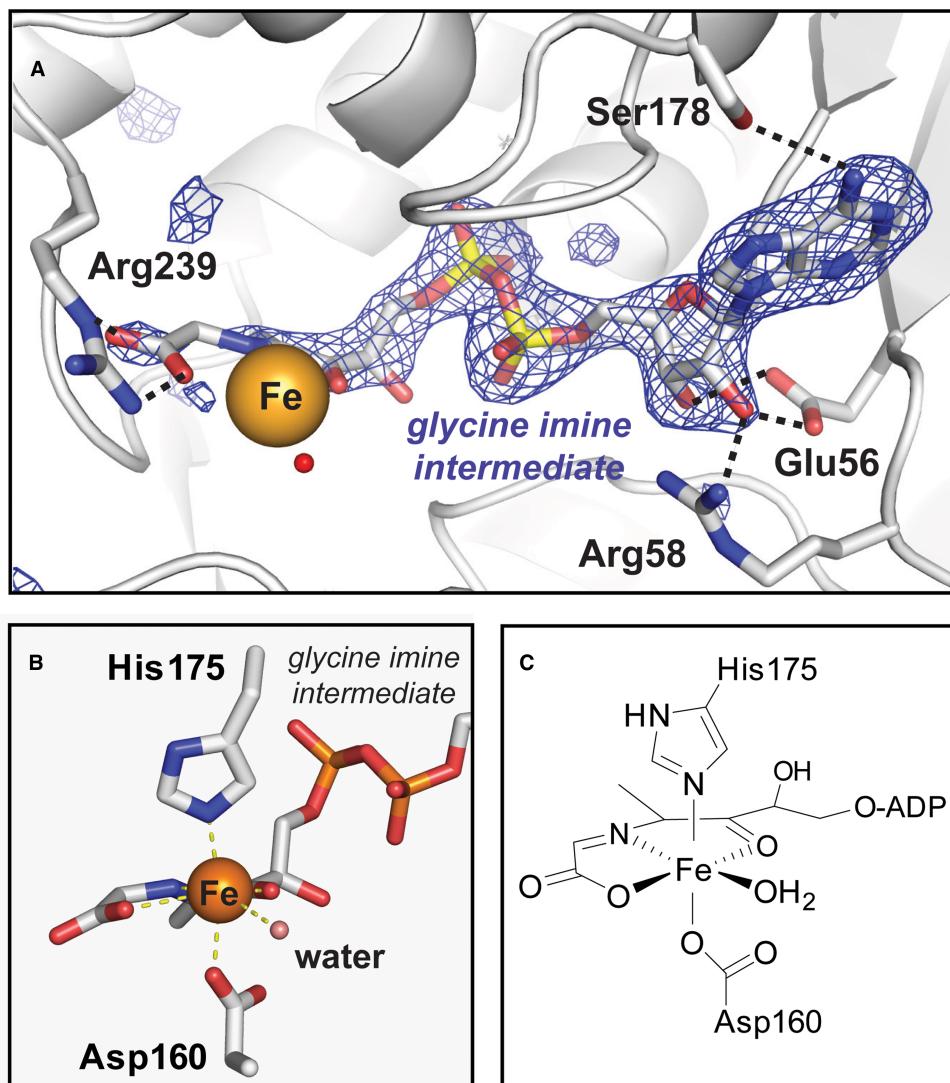


Figure 6. The active site of *Thermovibrio ammonificans* THI4.

(A) Interactions of the bound glycine intermediate with the TaThi4 protein active site. A glycine imine intermediate omit map, Fo–Fc contoured at  $3.5\sigma$ , is shown in blue. (B) Ligand geometry around the metal center with a near-octahedral coordination environment as diagrammed in (C).

Having established that certain non-Cys THI4s can — unexpectedly — operate in mild, fully aerobic conditions, we investigated characteristics that enable them to do so. We could dismiss any role, positive or negative, for thermophily because complementing activity varied similarly in thermophiles and mesophiles (Table 1). There was likewise no association between complementing activity and the  $O_2$  adaptation of the source organism (Table 1). Nor did the residue that replaces the sacrificial Cys or the number of Cys or Met residues appear to be important (Supplementary Table S3).

What then, might confer aerotolerance? The comparative genomic analysis provided a clue by associating non-Cys THI4s with proteins that bind, transport, or metabolize nonferrous transition metals, notably cobalt or nickel (Figure 3A), and the observed complementing activity of TaTHI4 in yeast favored the possibility that the metal cofactor is cobalt (Supplementary Figure S6). Neither the comparative genomics nor the yeast complementation data ruled out a ferrous iron cofactor, however.

Comparing the TaTHI4 structure — the first deposited for a bacterial THI4 — with those of the ecologically less  $O_2$ -exposed THI4s from archaeal methanogens (PDBs: 4Y4M, 4Y4N, 6HK1) [7,50] did not provide strong evidence on the aerotolerance phenomenon. The active sites are largely identical, with one major difference,

the variation of residues (Met vs. His) replacing the Cys residue of suicide THI4s. Variants at this position, however, do not correlate with aerotolerance (Supplementary Table S3). The octamer structures show similar dimensions of the overall barrel-like architecture and similar surface electrostatic charge. Notable differences are two extended loop regions in TaTHI4 at the entrance to the large active-site pore (residues 134–142 and 186–197); however, based on the pore diameter, these changes are unlikely to affect O<sub>2</sub> diffusion. An analogous comparision of the structures of O<sub>2</sub>-sensitive and O<sub>2</sub>-tolerant homologs of [NiFe] hydrogenase showed distinct, well-defined tunnels to the metal center [54,55]. In this case, structural variations of the tunnel were predicted to regulate O<sub>2</sub> diffusion to the reactive metal center. All THI4s have a large pore (~26 Å) linking the active site to solvent, precluding variation in O<sub>2</sub> diffusion. The lack of dissimilar structural features favors the alternative possibility of a role for the active site metal in O<sub>2</sub> tolerance.

Semi-quantitative *in vitro* assay of *M. jannaschii* THI4 activity [14] showed that cobalt supported ~60%, and nickel ~25%, of the activity given by iron. The presented TaTHI4 structure contains an obligate bound Fe(II) atom, based on Fe(II) being the only metal present during protein reconstitution. The ligand geometry in the structure (Figure 6) does not preclude other metals, including Co(II) or Ni(II). A bound Ni(II) would be predicted to favor a square pyramidal geometry [56] over the octahedral although the active site could possibly accommodate a minor shift in ligand position to square pyramidal. There is solid precedent for octahedral Co (II), including colbalamin enzymes and methylmalonyl-CoA carboxytransferase [57]. In terms of THI4 chemistry, Co(II) is less likely to react deleteriously with molecular O<sub>2</sub> than Ni(II) and Fe(II). Contrasting with many examples in which mononuclear Fe(II) reacts with O<sub>2</sub> to generate reactive O-atom transfer species in metalloenzymes and model compounds [58,59], reaction of Co(II) with O<sub>2</sub> has strong precedent for reversible formation of Co(III)-superoxide species in the absence of additional reducing equivalents or protonation [60], implying that replacing iron with cobalt could provide a mechanism for the observed O<sub>2</sub> tolerance. In addition, a Ni(II)-SH intermediate in the catalytic cycle would be particularly stable [61], disfavoring catalysis. Other metals are known to be able to perform THI4 chemistry, notably Zn(II), but the ligand geometry and metal preference analysis of *M. jannaschii* THI4 [14] argue against a tetrahedral liganded metal.

In summary, the comparative genomics, functional complementation, and structural evidence collectively implicate the bound metal as a natural determinant of THI4 aerotolerance, with Co(II) the best candidate although Ni(II) cannot be rigorously excluded. Definitive evidence on this point will require the development of quantitative *in vitro* assays for THI4 activity. Besides the nature of the metal inserted, there may well be other determinants of aerotolerance. We hope to identify such determinants, and to gain insight into the metal cofactor, from ongoing continuous directed evolution experiments [62] to improve the complementing activity of native non-Cys THI4s. If — as seems likely [63] — such improvement is possible, there is a realistic prospect of replacing suicidal plant THI4s with catalytic THI4s that work well in aerobic conditions and thus slash the energy cost of thiamin synthesis.

## Data Availability

Co-ordinates and structure factors of the *T. ammonificans* THI4 crystal structure have been deposited in PDB under code 7RK0.

## Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

## Funding

This work was supported primarily by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award DE-SC0020153, and by an endowment from the C.V. Griffin Sr. Foundation.

## Open Access

Open access for this article was enabled by the participation of University of Florida in an all-inclusive *Read & Publish* pilot with Portland Press and the Biochemical Society.

## Author Contribution

A.D.H., J.J., and S.D.B. designed the research; J.J., J.-D.G.-G., and B.J.L. performed assays; J.J. and A.D.H. ran bioinformatic analyses; Q.L., Y.H., and S.D.B. carried out structural analysis; A.D.H. wrote the article with input from J.J. and S.D.B.

## CRediT Author Contribution

**Andrew D. Hanson:** Conceptualization, formal analysis, supervision, funding acquisition, investigation, writing — original draft. **Jaya Joshi:** Conceptualization, formal analysis, investigation, writing — review and editing. **Qiang Li:** Formal analysis, investigation, writing — review and editing. **Jorge D. García-García:** Investigation, writing — review and editing. **Bryan J. Leong:** Investigation, writing — review and editing. **You Hu:** Investigation, writing — review and editing. **Steven D. Bruner:** Conceptualization, formal analysis, supervision, investigation, writing — original draft, writing — review and editing.

## Acknowledgements

We thank Prof. Leslie J. Murray for helpful discussions.

## Abbreviations

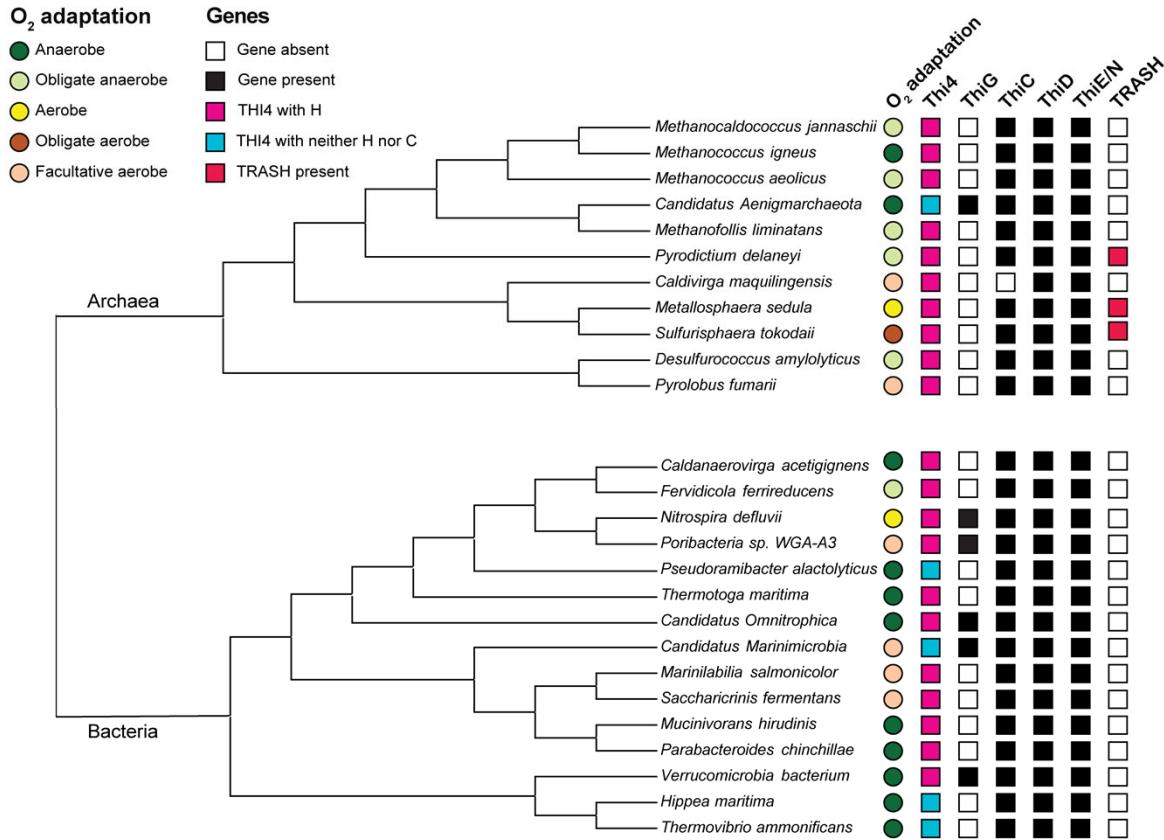
ADT, adenylated carboxythiazole; IPTG, Isopropyl  $\beta$ -D-1-thiogalactopyranoside; ROS, reactive oxygen species; SSN, sequence similarity network.

## References

- 1 Begley, T.P., Ealick, S.E. and McLafferty, F.W. (2012) Thiamin biosynthesis: still yielding fascinating biological chemistry. *Biochem. Soc. Trans.* **40**, 555–560 <https://doi.org/10.1042/BST20120084>
- 2 Hanson, A.D., Amthor, J.S., Sun, J., Niehaus, T.D., Gregory, III, J.F., Bruner, S.D. et al. (2018) Redesigning thiamin synthesis: prospects and potential payoffs. *Plant Sci.* **273**, 92–99 <https://doi.org/10.1016/j.plantsci.2018.01.019>
- 3 Chatterjee, A., Abeydeera, N.D., Bale, S., Pai, P.J., Dorrestein, P.C., Russell, D.H. et al. (2011) *Saccharomyces cerevisiae* THI4p is a suicide thiamine thiazole synthase. *Nature* **478**, 542–546 <https://doi.org/10.1038/nature10503>
- 4 Faou, P. and Tropschug, M. (2004) *Neurospora crassa* CyPBP37: a cytosolic stress protein that is able to replace yeast Thi4p function in the synthesis of vitamin B1. *J. Mol. Biol.* **344**, 1147–1157 <https://doi.org/10.1016/j.jmb.2004.09.097>
- 5 Joshi, J., Beaudoin, G.A.W., Patterson, J.A., García-García, J.D., Belisle, C.E., Chang, L.Y. et al. (2020) Bioinformatic and experimental evidence for suicidal and catalytic plant THI4s. *Biochem. J.* **477**, 2055–2069 <https://doi.org/10.1042/BCJ20200297>
- 6 Hwang, S., Cordova, B., Chavarria, N., Elbanna, D., McHugh, S., Rojas, J. et al. (2014) Conserved active site cysteine residue of archaeal THI4 homolog is essential for thiamine biosynthesis in *Halofexax volcanii*. *BMC Microbiol.* **14**, 260 <https://doi.org/10.1186/s12866-014-0260-0>
- 7 Zhang, X., Eser, B.E., Chanani, P.K., Begley, T.P. and Ealick, S.E. (2016) Structural basis for iron-mediated sulfur transfer in archaeal and yeast thiazole synthases. *Biochemistry* **55**, 1826–1838 <https://doi.org/10.1021/acs.biochem.6b00030>
- 8 Godoi, P.H., Galhardo, R.S., Luche, D.D., Van Sluys, M.A., Menck, C.F. and Oliva, G. (2006) Structure of the thiazole biosynthetic enzyme THI1 from *Arabidopsis thaliana*. *J. Biol. Chem.* **281**, 30957–30966 <https://doi.org/10.1074/jbc.M604469200>
- 9 Hanson, A.D., Beaudoin, G.A., McCarty, D.R. and Gregory, III, J.F. (2016) Does abiotic stress cause functional B vitamin deficiency in plants? *Plant Physiol.* **172**, 2082–2097 <https://doi.org/10.1104/pp.16.01371>
- 10 Nelson, C.J., Alexova, R., Jacoby, R.P. and Millar, A.H. (2014) Proteins with high turnover rate in barley leaves estimated by proteome analysis combined with in planta isotope labeling. *Plant Physiol.* **166**, 91–108 <https://doi.org/10.1104/pp.114.243014>
- 11 Li, L., Nelson, C.J., Trosch, J., Castleden, I., Huang, S. and Millar, A.H. (2017) Protein degradation rate in *Arabidopsis thaliana* leaf growth and development. *Plant Cell* **29**, 207–228 <https://doi.org/10.1105/tpc.16.00768>
- 12 Tivendale, N.D., Hanson, A.D., Henry, C.S., Hegeman, A.D. and Millar, A.H. (2020) Enzymes as parts in need of replacement – and how to extend their working life. *Trends Plant Sci.* **25**, 661–669 <https://doi.org/10.1016/j.tplants.2020.02.006>
- 13 Hanson, A.D., McCarty, D.R., Henry, C.S., Xian, X., Joshi, J., Patterson, J.A. et al. (2021) The number of catalytic cycles in an enzyme's lifetime and why it matters to metabolic engineering. *Proc. Natl Acad. Sci. U.S.A.* **118**, e2023348118 <https://doi.org/10.1073/pnas.2023348118>
- 14 Eser, B.E., Zhang, X., Chanani, P.K., Begley, T.P. and Ealick, S.E. (2016) From suicide enzyme to catalyst: the iron-dependent sulfide transfer in *Methanococcus jannaschii* thiamin thiazole biosynthesis. *J. Am. Chem. Soc.* **138**, 3639–3642 <https://doi.org/10.1021/jacs.6b00445>
- 15 Sun, J., Sigler, C.L., Beaudoin, G.A.W., Joshi, J., Patterson, J.A., Cho, K.H. et al. (2019) Parts-prospecting for a high-efficiency thiamin thiazole biosynthesis pathway. *Plant Physiol.* **179**, 958–968 <https://doi.org/10.1104/pp.18.01085>
- 16 Stewart, V. and Parales, Jr. J. (1988) Identification and expression of genes *narL* and *narX* of the *nar* (nitrate reductase) locus in *Escherichia coli* K-12. *J. Bacteriol.* **170**, 1589–1597 <https://doi.org/10.1128/jb.170.4.1589-1597.1988>
- 17 Neidhardt, F.C., Bloch, P.L. and Smith, D.F. (1974) Culture medium for enterobacteria. *J. Bacteriol.* **119**, 736–747 <https://doi.org/10.1128/jb.119.3.736-747.1974>
- 18 Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**, 5691–5702 <https://doi.org/10.1093/nar/gki866>
- 19 Suzek, B.E., Wang, Y., Huang, H., McGarvey, P.B. and Wu, C.H. (2015) Uniref clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 <https://doi.org/10.1093/bioinformatics/btu739>
- 20 Gerlt, J.A., Bouvier, J.T., Davidson, D.B., Imker, H.J., Sadkin, B., Slater, D.R. et al. (2015) Enzyme function initiative-enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* **1854**, 1019–1037 <https://doi.org/10.1016/j.bbapap.2015.04.015>
- 21 Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. U.S.A.* **97**, 6640–6645 <https://doi.org/10.1073/pnas.120163297>
- 22 Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M. et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 <https://doi.org/10.1038/msb4100050>

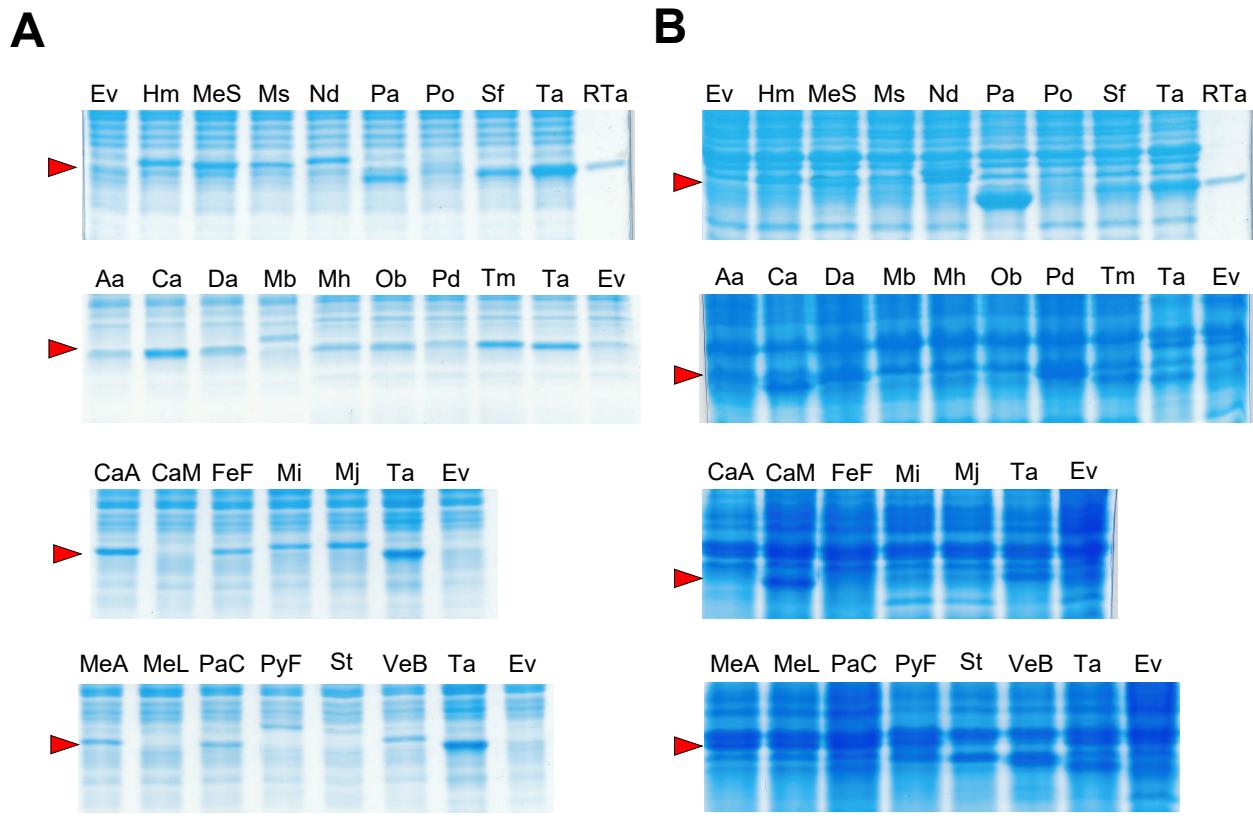
- 23 Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P<sub>BAD</sub> promoter. *J. Bacteriol.* **177**, 4121–4130 <https://doi.org/10.1128/jb.177.14.4121-4130.1995>
- 24 Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 <https://doi.org/10.1107/S0907444909047337>
- 25 Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M. et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 <https://doi.org/10.1107/S0907444912001308>
- 26 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 <https://doi.org/10.1107/S0907444904019158>
- 27 Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A. et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 <https://doi.org/10.1107/S0907444911001314>
- 28 Moriarty, N.W., Grosse-Kunstleve, R.W. and Adams, P.D. (2009) Electronic ligand builder and optimization workbench (elBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080 <https://doi.org/10.1107/S0907444909029436>
- 29 Zallot, R., Oberg, N. and Gerlt, J.A. (2019) The EFL web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* **58**, 4169–4182 <https://doi.org/10.1021/acs.biochem.9b00735>
- 30 Ettema, T.J., Huynen, M.A., de Vos, W.M. and van der Oost, J. (2003) TRASH: a novel metal-binding domain predicted to be involved in heavy-metal sensing, trafficking and resistance. *Trends Biochem. Sci.* **28**, 170–173 [https://doi.org/10.1016/S0968-0004\(03\)00037-9](https://doi.org/10.1016/S0968-0004(03)00037-9)
- 31 Monchy, S., Benotmane, M.A., Wattiez, R., van Aelst, S., Aquier, V., Borremans, B. et al. (2006) Transcriptomic and proteomic analyses of the pMOL30-encoded copper resistance in *Cupriavidus metallidurans* strain CH34. *Microbiology (Reading)* **152**, 1765–1776 <https://doi.org/10.1099/mic.0.28593-0>
- 32 Baker-Austin, C., Dopson, M., Wexler, M., Sawers, R.G. and Bond, P.L. (2005) Molecular insight into extreme copper resistance in the extremophilic archaeon '*Ferroplasma acidarmanus*' Fer1. *Microbiology (Reading)* **151**, 2637–2646 <https://doi.org/10.1099/mic.0.28076-0>
- 33 Boyd, E.S. and Barkay, T. (2012) The mercury resistance operon: from an origin in a geothermal environment to an efficient detoxification machine. *Front. Microbiol.* **3**, 349 <https://doi.org/10.3389/fmicb.2012.00349>
- 34 Szymanski, M.R., Fiebach, A.R., Tratschin, J.D., Gut, M., Ramanujam, V.M., Gottipati, K. et al. (2009) Zinc binding in pestivirus N(pro) is required for interferon regulatory factor 3 interaction and degradation. *J. Mol. Biol.* **391**, 438–449 <https://doi.org/10.1016/j.jmb.2009.06.040>
- 35 Orell, A., Remondellez, F., Arancibia, R. and Jerez, C.A. (2013) Molecular characterization of copper and cadmium resistance determinants in the biomining thermoacidophilic archaeon *Sulfolobus metallicus*. *Archaea* **2013**, 289236 <https://doi.org/10.1155/2013/289236>
- 36 Zhang, Y., Rodionov, D.A., Gelfand, M.S. and Gladyshev, V.N. (2009) Comparative genomic analyses of nickel, cobalt and vitamin B<sub>12</sub> utilization. *BMC Genomics* **10**, 78 <https://doi.org/10.1186/1471-2164-10-78>
- 37 Hausinger, R.P., Desguin, B., Fellner, M., Rankin, J.A. and Hu, J. (2018) Nickel-pincer nucleotide cofactor. *Curr. Opin. Chem. Biol.* **47**, 18–23 <https://doi.org/10.1016/j.cbpa.2018.06.019>
- 38 Schuelke-Sanchez, A.E., Stone, A.A. and Liptak, M.D. (2020) Cfba promotes insertion of cobalt and nickel into ruffled tetrapyrroles in vitro. *Dalton Trans.* **49**, 1065–1076 <https://doi.org/10.1039/C9DT03601F>
- 39 Guldan, H., Sterner, R. and Babinger, P. (2008) Identification and characterization of a bacterial glycerol-1-phosphate dehydrogenase: Ni<sup>2+</sup>-dependent AraM from *Bacillus subtilis*. *Biochemistry* **47**, 7376–7384 <https://doi.org/10.1021/bi8005779>
- 40 Korshunov, S., Imlay, K.R. and Imlay, J.A. (2016) The cytochrome bd oxidase of *Escherichia coli* prevents respiratory inhibition by endogenous and exogenous hydrogen sulfide. *Mol. Microbiol.* **101**, 62–77 <https://doi.org/10.1111/mmi.13372>
- 41 Hamza, M.A. and Engel, P.C. (2007) Enhancing long-term thermal stability in mesophilic glutamate dehydrogenase from *Clostridium symbiosum* by eliminating cysteine residues. *Enzyme Microb. Technol.* **41**, 706–710 <https://doi.org/10.1016/j.enzmotec.2007.06.008>
- 42 Yang, H., Liu, L., Li, J., Du, G. and Chen, J. (2012) Structure-based replacement of methionine residues at the catalytic domains with serine significantly improves the oxidative stability of alkaline amylase from alkaliphilic *Alkalimonas amylolytica*. *Biotechnol. Prog.* **28**, 1271–1277 <https://doi.org/10.1002/btpr.1611>
- 43 Hummel, C.S., Lancaster, K.M. and Crane, III, E.J. (2005) Determination of coenzyme A levels in *Pyrococcus furiosus* and other Archaea: implications for a general role for coenzyme A in thermophiles. *FEMS Microbiol. Lett.* **252**, 229–234 <https://doi.org/10.1016/j.femsle.2005.09.004>
- 44 Rocha, E.R., Tzianabos, A.O. and Smith, C.J. (2007) Thioredoxin reductase is essential for thiol/disulfide redox control and oxidative stress survival of the anaerobe *Bacteroides fragilis*. *J. Bacteriol.* **189**, 8015–8023 <https://doi.org/10.1128/JB.00714-07>
- 45 Heinemann, J., Hamerly, T., Maaty, W.S., Movahed, N., Steffens, J.D., Reeves, B.D. et al. (2014) Expanding the paradigm of thiol redox in the thermophilic root of life. *Biochim. Biophys. Acta* **1840**, 80–85 <https://doi.org/10.1016/j.bbagen.2013.08.009>
- 46 Herrmann, J.M. and Jakob, U. (2008) Special issue: redox regulation of protein folding. *Biochim. Biophys. Acta* **1783**, 519 <https://doi.org/10.1016/j.bbampcr.2008.03.002>
- 47 Leary, S.C. (2010) Redox regulation of SCO protein function: controlling copper at a mitochondrial crossroad. *Antioxid. Redox Signal.* **13**, 1403–1416 <https://doi.org/10.1089/ars.2010.3116>
- 48 Vetriani, C., Speck, M.D., Ellor, S.V., Lutz, R.A. and Starovoytov, V. (2004) *Thermovibrio ammonificans* sp. nov., a thermophilic, chemolithotrophic, nitrate-ammonifying bacterium from deep-sea hydrothermal vents. *Int. J. Syst. Evol. Microbiol.* **54**, 175–181 <https://doi.org/10.1099/ijs.0.02781-0>
- 49 Giovannelli, D., Sievert, S.M., Hügler, M., Markert, S., Becher, D., Schweder, T. et al. (2017) Insight into the evolution of microbial metabolism from the deep-branching bacterium, *Thermovibrio ammonificans*. *eLife* **6**, e18990 <https://doi.org/10.7554/eLife.18990>
- 50 Engilberge, S., Wagner, T., Santoni, G., Breyton, C., Shima, S., Franzetti, B. et al. (2019) Protein crystal structure determination with the crystallophore, a nucleating and phasing agent. *J. Appl. Crystallogr.* **52**, 722–731 <https://doi.org/10.1107/S1600576719006381>
- 51 Stewart, L.C., Algar, C.K., Fortunato, C.S., Larson, B.I., Vallino, J.J., Huber, J.A. et al. (2019) Fluid geochemistry, local hydrology, and metabolic activity define methanogen community size and composition in deep-sea hydrothermal vents. *ISME J.* **13**, 1711–1721 <https://doi.org/10.1038/s41396-019-0382-3>
- 52 Pimentel, C., Caetano, S.M., Menezes, R., Figueira, I., Santos, C.N., Ferreira, R.B. et al. (2014) Yap1 mediates tolerance to cobalt toxicity in the yeast *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* **1840**, 1977–1986 <https://doi.org/10.1016/j.bbagen.2014.01.032>
- 53 Milne, N., Luttki, M.A.H., Cueto Rojas, H.F., Wahl, A., van Maris, A.J.A., Pronk, J.T. et al. (2015) Functional expression of a heterologous nickel-dependent, ATP-independent urease in *Saccharomyces cerevisiae*. *Metab. Eng.* **30**, 130–140 <https://doi.org/10.1016/j.ymben.2015.05.003>

- 54 Kalms, J., Schmidt, A., Frielingsdorf, S., van der Linden, P., von Stetten, D., Lenz, O. et al. (2016) Krypton derivatization of an O<sub>2</sub>-tolerant membrane-bound [NiFe] hydrogenase reveals a hydrophobic tunnel network for gas transport. *Angew. Chem. Int. Ed. Engl.* **55**, 5586–5590 <https://doi.org/10.1002/anie.201508976>
- 55 Kalms, J., Schmidt, A., Frielingsdorf, S., Utesch, T., Gotthard, G., von Stetten, D. et al. (2018) Tracking the route of molecular oxygen in O<sub>2</sub>-tolerant membrane-bound [NiFe] hydrogenase. *Proc. Natl Acad. Sci. U.S.A.* **115**, E2229–E2237 <https://doi.org/10.1073/pnas.1712267115>
- 56 Alfano, M. and Cavazza, C. (2020) Structure, function, and biosynthesis of nickel-dependent enzymes. *Protein Sci.* **29**, 1071–1089 <https://doi.org/10.1002/pro.3836>
- 57 Kobayashi, M. and Shimizu, S. (1999) Cobalt proteins. *Eur. J. Biochem.* **261**, 1–9 <https://doi.org/10.1046/j.1432-1327.1999.00186.x>
- 58 Fukuzumi, S., Cho, K.B., Lee, Y.M., Hong, S. and Nam, W. (2020) Mechanistic dichotomies in redox reactions of mononuclear metal-oxygen intermediates. *Chem. Soc. Rev.* **49**, 8988–9027 <https://doi.org/10.1039/D0CS01251C>
- 59 Wang, Y., Li, J. and Liu, A. (2017) Oxygen activation by mononuclear nonheme iron dioxygenases involved in the degradation of aromatics. *J. Biol. Inorg. Chem.* **22**, 395–405 <https://doi.org/10.1007/s00775-017-1436-5>
- 60 Fiedler, A.T. and Fischer, A.A. (2017) Oxygen activation by mononuclear Mn, Co, and Ni centers in biology and synthetic complexes. *J. Biol. Inorg. Chem.* **22**, 407–424 <https://doi.org/10.1007/s00775-016-1402-7>
- 61 Shearer, J., Peck, K.L., Schmitt, J.C. and Neupane, K.P. (2014) Cysteinate protonation and water hydrogen bonding at the active-site of a nickel superoxide dismutase metallopeptide-based mimic: implications for the mechanism of superoxide reduction. *J. Am. Chem. Soc.* **136**, 16009–16022 <https://doi.org/10.1021/ja5079514>
- 62 García-García, J.D., Joshi, J., Patterson, J.A., Trujillo-Rodriguez, L., Reisch, C.R., Javanpour, A.A. et al. (2020) Potential for applying continuous directed evolution to plant enzymes: an exploratory study. *Life (Basel)* **10**, 179 <https://doi.org/10.3390/life10090179>
- 63 Rix, G. and Liu, C.C. (2021) Systems for in vivo hypermutation: a quest for scale and depth in directed evolution. *Curr. Opin. Chem. Biol.* **64**, 20–26 <https://doi.org/10.1016/j.cbpa.2021.02.008>



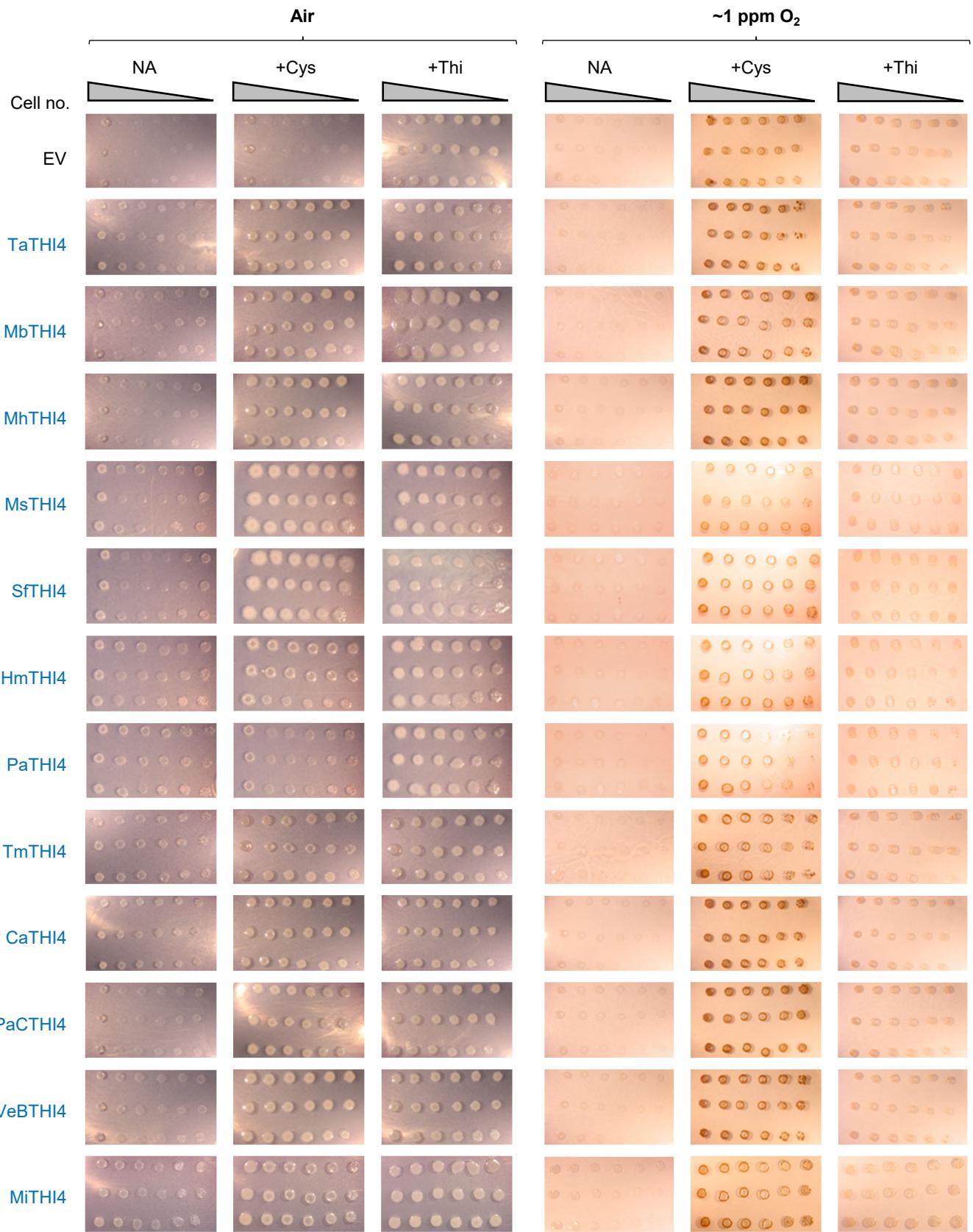
**Supplementary Figure 1. Ecology and genomic context of the 26 THI4s selected for testing.**

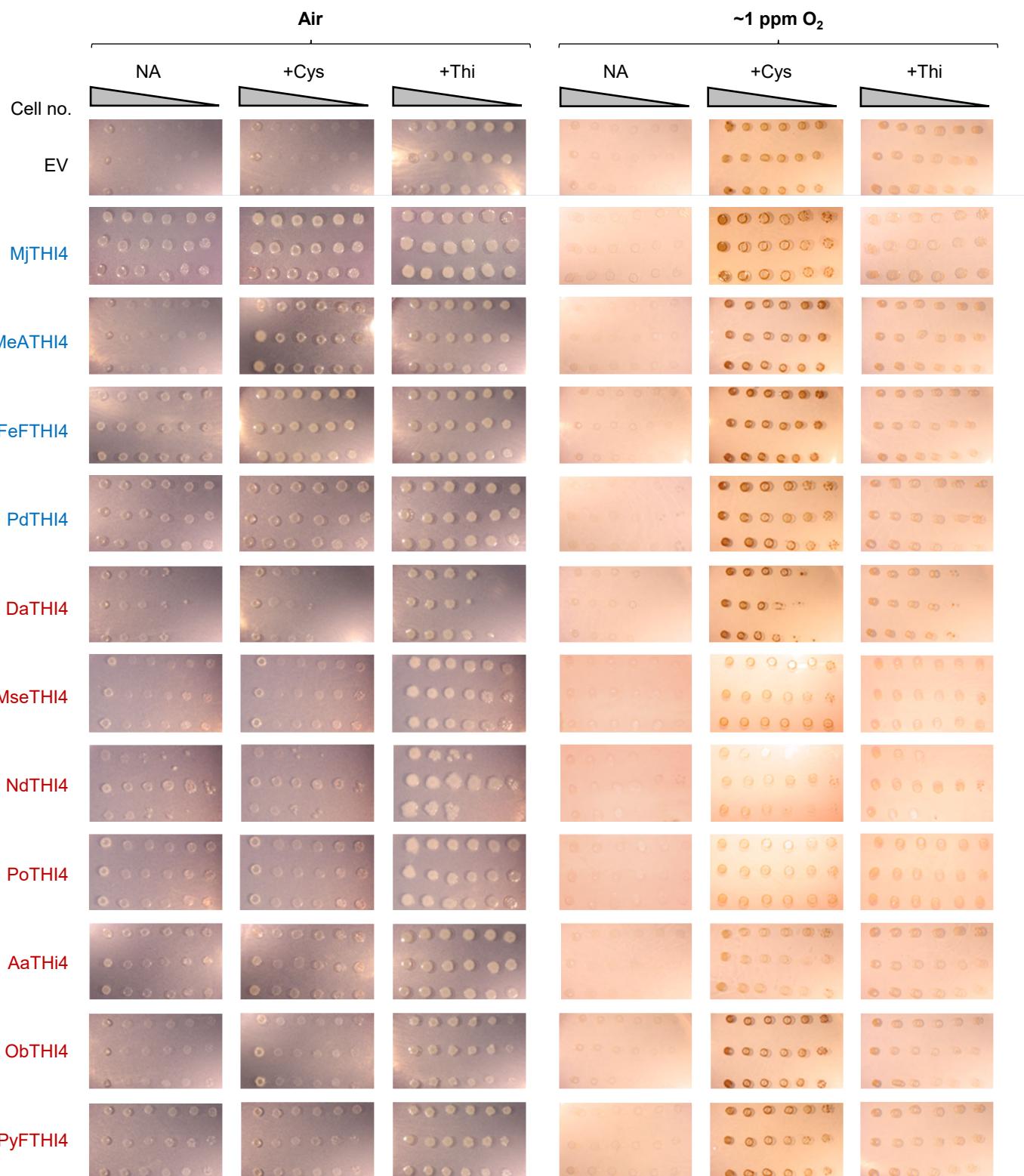
The first column (colored bullets) shows the O<sub>2</sub> adaptation of the 15 bacteria and 11 archaea whose THI4s were tested. The second column (colored squares) shows which residue replaces the active-site cysteine in each THI4. The next four columns (black or white squares) indicate the presence or absence of other thiamin synthesis enzymes (ThiG, ThiC, ThiD, and ThiE or ThiN). The last column (red or white squares) indicates presence or absence of a gene encoding a protein from the TRASH family (Trafficking, Resistance, And Sensing of Heavy metals) that is clustered with the THI4 gene.



**Supplementary Figure 2. Soluble expression of non-Cys THI4s.**

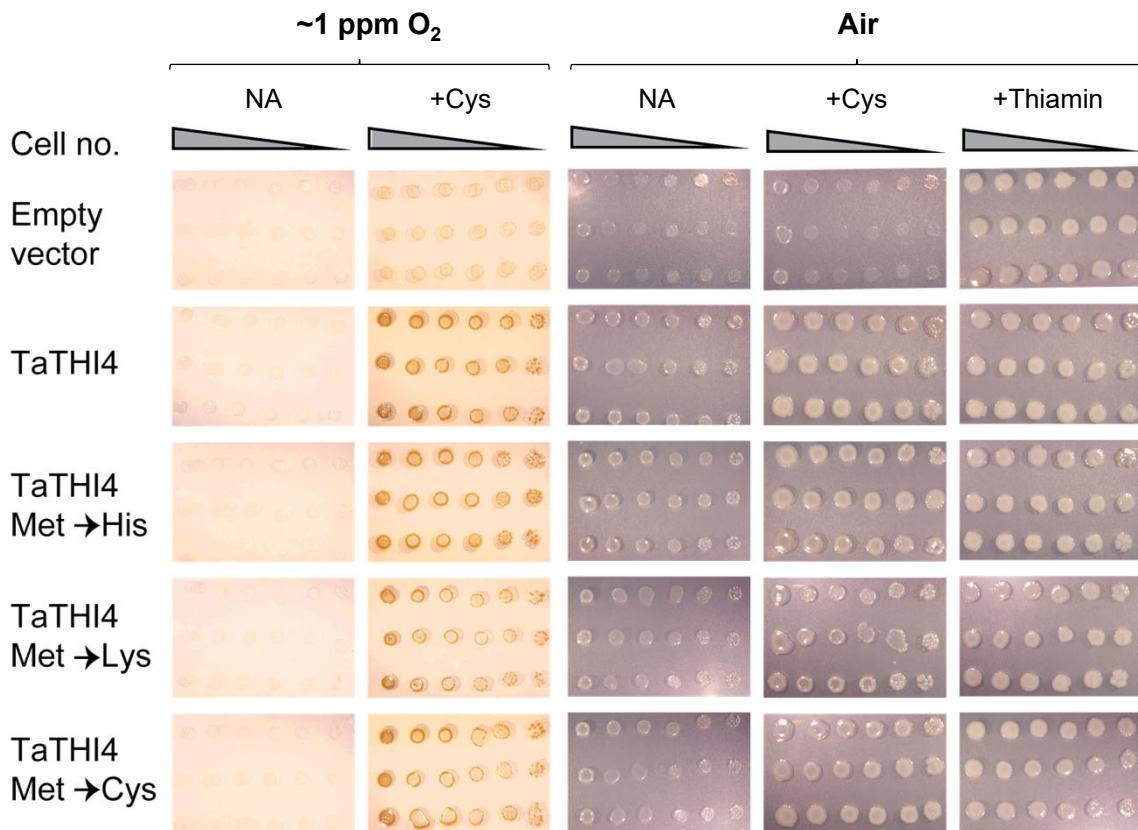
Quantitative gel analysis of (**A**) soluble and (**B**) insoluble expression in *E. coli* of 26 selected non-Cys THI4s. Soluble and insoluble fractions of cells were run on 15% gels, stained with Coomassie blue, and scanned to quantify the THI4 band, for which purified recombinant *Thermovibrio ammonificans* THI4 (RTa, arrow) served as a marker. Organism abbreviations are as in Table 1.





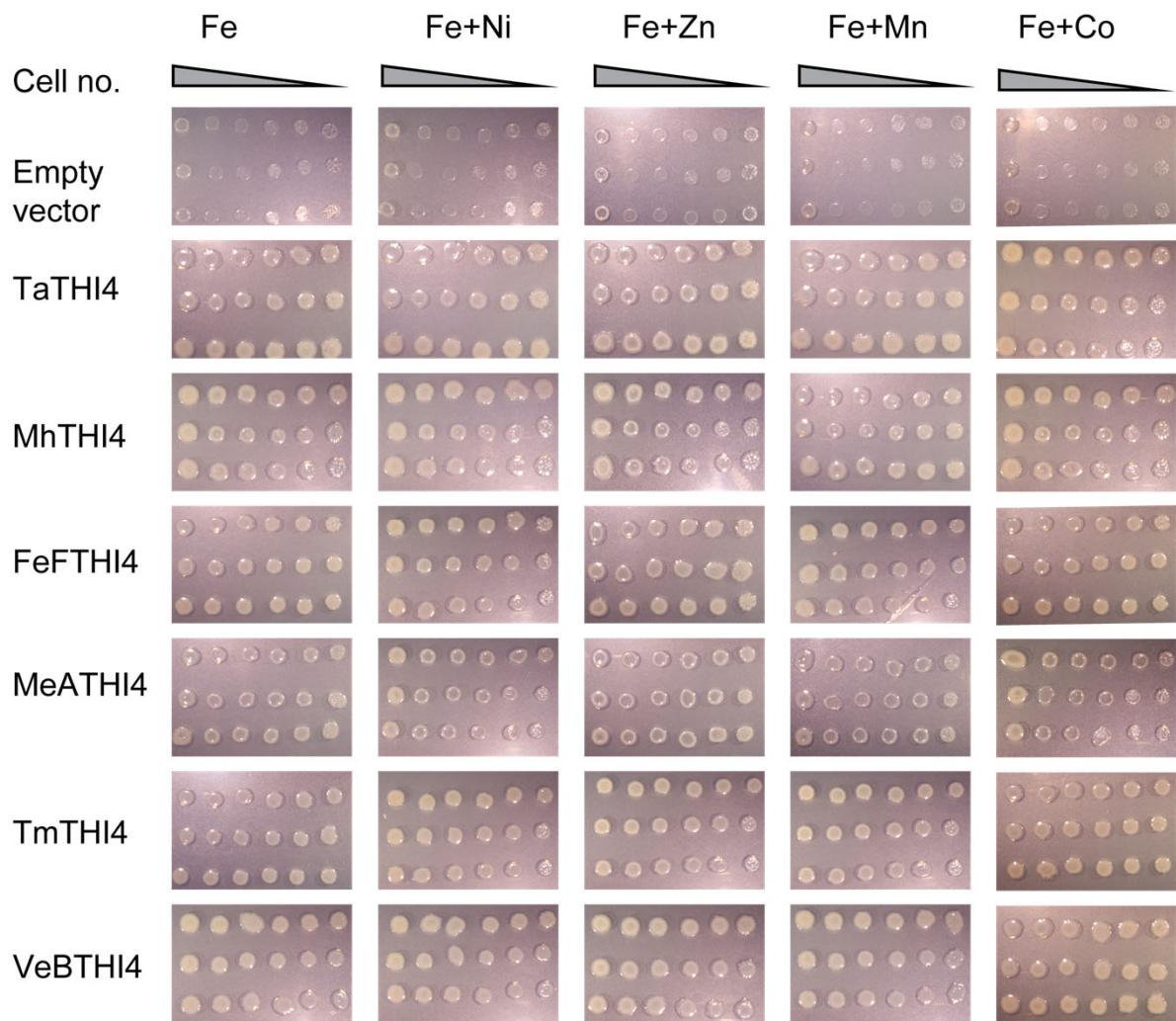
**Supplementary Figure 3. Functional complementation tests of non-Cys THI4s.**

Tests of functional complementation of an *E. coli*  $\Delta\text{thiG}$  strain by all 23 soluble non-Cys THI4s or the empty vector (EV). Organism abbreviations are as in Table 1. Overnight cultures of three independent clones per construct were 10-fold serially diluted and spotted on plates of MOPS minimal medium containing 0.2% glycerol and 0.02% arabinose with no additions (NA) or plus 1 mM Cys or 100 nM thiamin. Cells were cultured in air or  $\sim 1 \text{ ppm O}_2$ . The medium used for culture in  $\sim 1 \text{ ppm O}_2$  contained 40 mM nitrate. Images were captured after incubation at 37°C for 7 d. The high background in the  $\sim 1 \text{ ppm O}_2$ +Cys treatment is staining of the inoculum cells. Organisms whose THI4s showed clear complementing activity in air, particularly with Cys supplementation, are blue; organisms whose THI4s did not show such activity are red. Note that complementing activity was scored from direct visual inspection of plates, not from the above images, which do not fully capture growth in every case.



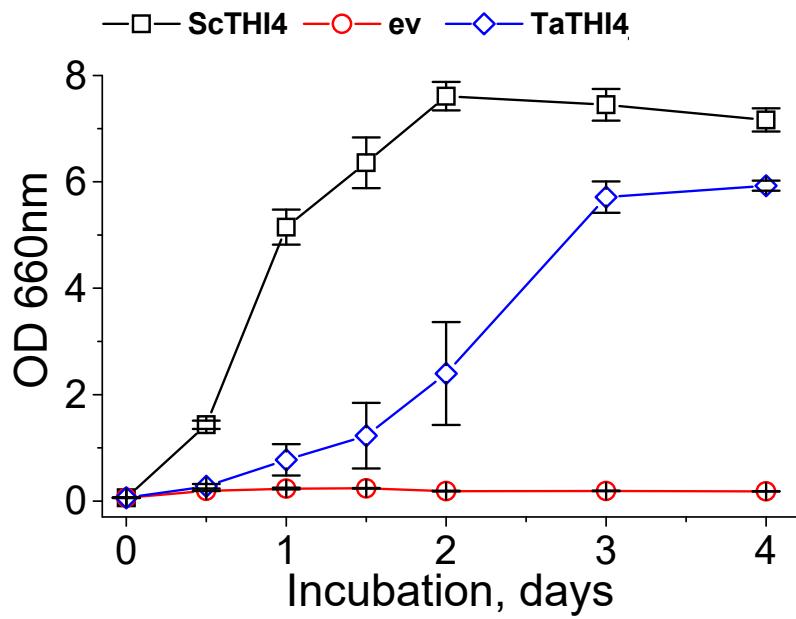
**Supplementary Figure 4. Complementation activity of TaTHI4 mutants**

An *E. coli*  $\Delta thiG$  strain was transformed with empty vector or vector harboring wild type TaTHI4 or TaTHI4 with the indicated mutations of Met158. Overnight cultures of three independent clones per construct were 10-fold serially diluted and spotted on plates of MOPS minimal medium containing 0.2% (w/v) glycerol and 0.02% (w/v) arabinose with no addition (NA) or with 1 mM Cys or 100 nM thiamin. The medium used for culture in ~1 ppm O<sub>2</sub> contained 40 mM nitrate. Cells were cultured in air or under N<sub>2</sub> containing ~1 ppm O<sub>2</sub>. Images were captured after incubation at 37°C for 7 d.



**Supplementary Figure 5 Effect of metal supplementation on complementing activity of THI4s**

An *E. coli*  $\Delta thiG$  strain was transformed with empty vector or vector harboring the indicated THI4 sequence. Overnight cultures of three independent clones per construct were 10-fold serially diluted and spotted on plates of MOPS minimal medium containing 0.2% (w/v) glycerol, 0.02% (w/v) arabinose, 1 mM Cys, and 100  $\mu$ M of the indicated metal. All media also contained the standard concentration of ferrous iron (100  $\mu$ M). Cultures were incubated in air at 37°C for 7 d.



**Supplementary Figure 6 Complementation of a yeast  $\Delta THI4$  strain by TaTHI4**

Cells of the  $\Delta THI4$  strain transformed with empty vector (ev) or vector harboring TaTHI4 or yeast THI4 (ScTHI4) as positive control were cultured in thiamin-free SC minus histidine medium. Data are means of three or four independent clones  $\pm$  S.E.

**Supplementary Table 2** *T. ammonificans* THI4 data collection and refinement statistics

|                                     | TaThi4                  |
|-------------------------------------|-------------------------|
| <b>Data collection</b>              |                         |
| Space group                         | I121                    |
| Cell dimensions                     |                         |
| $a, b, c$ (Å)                       | 89.89, 89.69, 131.81    |
| $\alpha, \beta, \gamma$ (°)         | 90, 96.98, 90           |
| Resolution (Å)                      | 43.63-2.30 (2.36-2.30)* |
| $R_{\text{merge}}$                  | 0.092 (1.552)           |
| $I / \sigma I$                      | 8.23 (0.77)             |
| Completeness (%)                    | 99.46 (96.62)           |
| Redundancy                          | 2.0 (2.0)               |
| <b>Refinement</b>                   |                         |
| Resolution (Å)                      | 43.63-2.30              |
| No. reflections                     | 93685                   |
| $R_{\text{work}} / R_{\text{free}}$ | 0.217/0.275             |
| No. atoms                           |                         |
| Protein                             | 7835                    |
| Ligand/ion                          | 160                     |
| Water                               | 58                      |
| $B$ -factors                        |                         |
| Protein                             | 50.69                   |
| Ligand/ion                          | 41.32                   |
| Water                               | 47.34                   |
| R.m.s. deviations                   |                         |
| Bond lengths (Å)                    | 0.009                   |
| Bond angles (°)                     | 1.12                    |

\*Values in parentheses are for highest-resolution shell.

### Supplementary Table 3 Sequences of 199 non-Cys THI4s from SEED and UniRef90 databases

The 26 sequences selected for testing are boxed. Those that had little to no complementing activity are in **red font**. Those that were poorly expressed in soluble form in *E. coli* and were not tested are in **gray font**. The residue that replaces Cys in the active site is highlighted in **cyan**. Met and non-active-site Cys residues in each sequence are highlighted in **yellow**; the number per sequence is given in the header. Mean values were: Active THI4s: 10.6 Met, 2.3 Cys. Inactive THI4s: 9.1 Met, 3.1 Cys.

|   |
|---|
| >Caldanaerovirga acetigignens (2 Cys, 11 Met)   |
| MKSFSIPVPTDKVSSLIMKHYFKDLEDAVKSDVIVAGAGPSGLTC <sup>A</sup> WTLADQGYKVTLDRLAPGGIWGGAMSFNKVVLQKDVEWILKEADVPFV<br>EDEGALVVSAPIFLASKLIAAKAAAHPGIRFN <sup>M</sup> LTVVDLHSSGDRITGVVVNNSAIE <sup>M</sup> AGLH <sup>V</sup> DPMVLTAKVLATGHDAVLANLYSRRAGTGLIRE<br>SFMNAEKGEEDVVANTRMLAPGLFVAGMAANNVEGGCRMGP <sup>I</sup> FGGMLSGKKAARLIINYLSSENSK   |
| >Candidatus Marinimicrobia bacterium (4 Cys, 12 Met)  |
| MEKIVSFGIIDSYQKKLKENLEVDVAIVGGGPSGLIAAKYLAQAGKKVVLFERKLAPGGGMWGGAMMFNFNQIVVQEDAISILEDVDISYNLYEEGYYV<br>CD <sup>S</sup> VEATAALI <sup>S</sup> AKKAGATIFN <sup>C</sup> SVEDVV <sup>F</sup> QRG <sup>S</sup> VAGVVNVASVHREG <sup>M</sup> YVDP <sup>L</sup> VIMAKA <sup>V</sup> LDSTGH <sup>S</sup> CEVASILARKNEV <sup>K</sup> LM <sup>T</sup> R <sup>R</sup> G <sup>N</sup> IM <sup>G</sup> ER<br>SLSIEEGELTTIENTKEIFPGLYVSGMAANAVSGSFRMGP <sup>I</sup> FGGMLSGKKAARLIINYLSSENSK   |
| >Fervidicola ferrireducens (2 Cys, 11 Met)  |
| MKSFSVPVPTDKVSSLIMKHYFKDLEDAVKSDVIVAGAGPSGLTC <sup>A</sup> WTLADQGYKVTLDRLAPGGIWGGAMSFNKVVLQKDVEWILKEADVPFV<br>EDEGALVVSAPIFLASKLIAAKAAAHPGIRFN <sup>M</sup> LTVVDLHSTGD <sup>K</sup> ITGVVVNNSAIE <sup>M</sup> AGLH <sup>V</sup> DPMVLTAKVLATGHDAVLANLYSRRAGTGLIRE<br>SFMNAEKGEEDVVANTRMLAPGLFVAGMAANNVEGGCRMGP <sup>I</sup> FGGMLSGKKAARLIINYLSSENSK  |
| >Hippea maritima strain ATCC 700847 (2 Cys, 9 Met)  |
| MNNLDERVISRAI <sup>V</sup> ERYMNKLLDY <sup>C</sup> D <sup>T</sup> IVGGGPAGLV <sup>C</sup> AYYLAKANIKVAIFDKRLTIGGGMWGGAMLFNEIVVQEIGREILDEFGINYEKYTD<br>GYTADSIEATT <sup>T</sup> LISTKVAGAKIFN <sup>C</sup> FAI <sup>E</sup> VEDVV <sup>F</sup> KKIDGQYR <sup>V</sup> NGLVVGWT <sup>V</sup> NNAGL <sup>H</sup> IVDPLV <sup>T</sup> SKYVIDATGH <sup>D</sup> ADIANILTR <sup>K</sup> GGIKLNTP<br>EGVVIGEKPMWAEV <sup>G</sup> EQ <sup>S</sup> TIEETQEV <sup>V</sup> PG <sup>L</sup> IVAC <sup>M</sup> AAAVAVSGSHRMGP <sup>I</sup> FGGMLNSGKKAQIVIESLK   |
| >Marinilabilia salmonicolor (2 Cys, 12 Met)   |
| MEQIVSSGIIDSYFSKLKENLA <sup>V</sup> DAIVGGGPSGLIA <sup>A</sup> YYLA <sup>T</sup> K <sup>G</sup> KVALFERKLAPGGGMWGGAMMFNEIMVQKEALHILKELGIEYKHYRDDYT<br>VDSVHATSALTYHATKAGARIFN <sup>C</sup> T <sup>S</sup> I <sup>E</sup> EDVV <sup>F</sup> HNNIVSGL <sup>V</sup> INWAPVHREG <sup>M</sup> HVDPLI <sup>I</sup> MAKAVIDGTGH <sup>D</sup> CEIVHTVARKNDIKIDTPSGKVM <sup>G</sup> ER<br>SLAVEEAERTTVDNTKEV <sup>P</sup> GLFVSGMAANGTGSYRMGP <sup>I</sup> FGGMLLSGQKVAGI <sup>S</sup> E <sup>K</sup> LAKAEIMESANN   |
| >Methanocaldococcus jannaschii DSM 2661 (= Methanococcus jannaschii) (3 Cys, 10 Met)  |
| MVNLMNIKDIKLN <sup>A</sup> DET <sup>T</sup> KT <sup>K</sup> AILKASFD <sup>M</sup> WLDIVEADVV <sup>V</sup> GAGPSGLTC <sup>A</sup> RYLAKEGFKVVVLERHLAFGGTWGGGMGF <sup>P</sup> YIVVEEPADELLRE<br>VG <sup>I</sup> KLIDMGDGYYVADSVEVPAKLA <sup>V</sup> AA <sup>M</sup> DAGAKILT <sup>G</sup> IVVEDL <sup>L</sup> REDGVAGVVINSYAIERAGL <sup>H</sup> IDPLT <sup>I</sup> R <sup>S</sup> KVV <sup>V</sup> DATGHEASIVN <sup>I</sup> LVKKNK<br>LEADVPGEKSMWA <sup>E</sup> KGENALLRN <sup>T</sup> REVY <sup>P</sup> GLFVSG <sup>M</sup> AA <sup>N</sup> ASHGGYRMGAIFGGMYLSGKLC <sup>A</sup> ELITE <sup>K</sup> LKN  |
| >Methanotorris igneus (= Methanococcus igneus) (2 Cys, 8 Met)   |
| MDVRLRADEYATTRAILKSAFD <sup>M</sup> WLDI <sup>I</sup> DVDVAIVGGGPSGLTA <sup>R</sup> YIAKEGYKV <sup>V</sup> LERHLAFGGTWGGGMGF <sup>P</sup> YIVVEEPADEILREVGV <sup>K</sup> LEK<br>VEGEDGLY <sup>T</sup> ADSVEVPAKLA <sup>V</sup> GA <sup>I</sup> DAGAKVL <sup>T</sup> GIVVEDL <sup>V</sup> RENRVAGVVINSYAI <sup>E</sup> KAGL <sup>H</sup> IDP <sup>I</sup> ITAKYVV <sup>D</sup> ATGH <sup>D</sup> ASV <sup>T</sup> TSRK <sup>N</sup> PEL <sup>G</sup> LE<br>VPGEKSMWA <sup>E</sup> KGENALLRN <sup>T</sup> REVY <sup>P</sup> GLFVSG <sup>M</sup> AA <sup>N</sup> AVYAGHRMGAIFGGMYISGK <sup>K</sup> CAEM <sup>I</sup> VE <sup>K</sup> LKN   |
| >Methanococcus aeolicus strain ATCC BAA-1280 (3 Cys, 8 Met)   |
| MDISKIDLKADEKAVTKSIFKATYEMW <sup>D</sup> NLEVDVV <sup>V</sup> VGGGPSGLTAGRYLADAGVKVL <sup>L</sup> ILERHLSFGGGTWGGGMCP <sup>I</sup> TYVQSPADEIILSEVG <sup>I</sup> K<br>LEE <sup>E</sup> GEDGLFVADSVEVPAKLG <sup>T</sup> GA <sup>I</sup> DAGAKVL <sup>T</sup> GIVVEDV <sup>L</sup> KEGV <sup>S</sup> V <sup>V</sup> INSYAI <sup>K</sup> AGL <sup>H</sup> IDPLT <sup>I</sup> NA <sup>K</sup> YVIDATGH <sup>D</sup> ASV <sup>T</sup> CT <sup>L</sup> ARKNED <sup>I</sup> G <sup>L</sup><br>VIPGEKSLWA <sup>D</sup> E <sup>G</sup> EN <sup>G</sup> LL <sup>K</sup> Y <sup>T</sup> KEL <sup>F</sup> GLFV <sup>C</sup> GMASNATHGGYRMGA <sup>V</sup> FGGMYISGKIVADMILEKLKN  |
| >Mucinivorans hirudinis (3 Cys, 10 Met)   |
| MEKIVSAGIVESYFD <sup>K</sup> LRRNLVLD <sup>A</sup> IVGGGPSGLVAA <sup>Y</sup> YLA <sup>T</sup> KAGR <sup>R</sup> VALFERKLAPGGGMWGGAMMFNDIVVQSDALPILEELGVSYRH <sup>R</sup> DAYL<br>VDSVHATA <sup>A</sup> LYA <sup>T</sup> RA <sup>G</sup> ATIFN <sup>C</sup> YSVEDVV <sup>F</sup> K <sup>D</sup> ERVAGL <sup>V</sup> VNW <sup>A</sup> PI <sup>R</sup> EG <sup>M</sup> HVDPL <sup>I</sup> V <sup>M</sup> ATA <sup>V</sup> LEG <sup>T</sup> GH <sup>D</sup> CA <sup>I</sup> ARL <sup>V</sup> ARKNGVR <sup>L</sup> NT <sup>P</sup> GEV <sup>I</sup> GER<br>SLSIEEAERTTVENTKEI <sup>Y</sup> PG <sup>L</sup> FLFVSG <sup>M</sup> AA <sup>N</sup> GV <sup>S</sup> GSFRMGP <sup>I</sup> FGGMLSGK <sup>K</sup> AAQ <sup>M</sup> ICDSL   |
| >Parabacteroides chinchillae (3 Cys, 11 Met)  |
| MEQIVSTGIIDSYFAKL <sup>K</sup> NSNL <sup>S</sup> VD <sup>A</sup> IVGGGPSGLVAA <sup>Y</sup> YLA <sup>T</sup> KAGR <sup>K</sup> VALFDRKLAPGGGMWGGAMMFNDIVVQSDALPILEELGVSYHAAGNG <sup>T</sup> YI<br>MDSVHTTSALIYQATKAGATIFN <sup>C</sup> YSVEDVV <sup>F</sup> H <sup>D</sup> NA <sup>V</sup> AGVV <sup>V</sup> NW <sup>A</sup> PI <sup>R</sup> EG <sup>M</sup> HVDPL <sup>I</sup> MA <sup>K</sup> AV <sup>L</sup> EG <sup>T</sup> GH <sup>D</sup> CE <sup>V</sup> ARTVARKNDIKLNT <sup>P</sup> GG <sup>I</sup> GER<br>SLNVELGESTTVENTKEI <sup>Y</sup> PG <sup>L</sup> FLFVSG <sup>M</sup> AA <sup>N</sup> GV <sup>S</sup> GSFRMGP <sup>I</sup> FGGMLSGK <sup>K</sup> AAELIC <sup>D</sup> DKLGK  |
| >Pseudoramibacter alactolyticus ATCC 23263 (2 Cys, 19 Met)  |
| M <sup>L</sup> SDTKISEA <sup>I</sup> LT <sup>T</sup> Y <sup>T</sup> DRFKQM <sup>L</sup> SSDA <sup>V</sup> IVGGGPSGLIA <sup>A</sup> YYLGKAGVK <sup>T</sup> LLDRRL <sup>S</sup> V <sup>G</sup> GGMM <sup>M</sup> QIVVQKS <sup>V</sup> L <sup>P</sup> ILEEM <sup>M</sup> GIACKYDAEH<br>YT <sup>V</sup> SSVAC <sup>I</sup> SGLIFRAAQSGATIMNL <sup>V</sup> T <sup>M</sup> EDAVV <sup>R</sup> EG <sup>L</sup> V <sup>I</sup> N <sup>W</sup> ST <sup>V</sup> TE <sup>M</sup> AHLM <sup>V</sup> DPL <sup>M</sup> MDARV <sup>V</sup> L <sup>D</sup> ATGH <sup>D</sup> AA <sup>L</sup> V <sup>T</sup> KL <sup>V</sup> ER <sup>M</sup> GP <sup>V</sup> L <sup>N</sup> TPSG <sup>G</sup> LEG<br>E <sup>K</sup> PKM <sup>W</sup> ADHG <sup>E</sup> KQV <sup>V</sup> ANTREVY <sup>P</sup> GLYVSG <sup>M</sup> AA <sup>N</sup> ATFG <sup>Q</sup> RMGP <sup>V</sup> FGGMLSGK <sup>K</sup> AAE <sup>M</sup> LR <sup>R</sup> LAQ  |
| >Pyrodictium delaneyi (0 Cys, 10 Met)   |
| M <sup>G</sup> IASF <sup>F</sup> Y <sup>P</sup> GE <sup>E</sup> L <sup>K</sup> QY <sup>S</sup> EA <sup>K</sup> L <sup>A</sup> IK <sup>V</sup> ALE <sup>K</sup> LSY <sup>A</sup> VE <sup>V</sup> AD <sup>V</sup> AIAGAG <sup>P</sup> AG <sup>L</sup> T <sup>L</sup> AW <sup>L</sup> LA <sup>E</sup> Q <sup>G</sup> LR <sup>V</sup> T <sup>L</sup> VE <sup>H</sup> RL <sup>S</sup> T <sup>GG</sup> M <sup>K</sup> GG <sup>S</sup> M <sup>L</sup> F <sup>P</sup> VAL <sup>V</sup> E <sup>E</sup> CLAA <sup>V</sup> IL<br>E <sup>K</sup> AG <sup>V</sup> RL <sup>H</sup> RV <sup>G</sup> E <sup>G</sup> LYAM <sup>D</sup> P <sup>V</sup> EA <sup>V</sup> AK <sup>L</sup> T <sup>A</sup> RA <sup>V</sup> DAG <sup>A</sup> V <sup>I</sup> L <sup>P</sup> GL <sup>H</sup> VE <sup>D</sup> LI <sup>V</sup> R <sup>G</sup> SG <sup>S</sup> N <sup>V</sup> R <sup>V</sup> AG <sup>I</sup> V <sup>V</sup> N <sup>W</sup> AP <sup>V</sup> VE <sup>A</sup> GW <sup>H</sup> V <sup>D</sup> PL <sup>I</sup> Y <sup>E</sup> AR <sup>A</sup> V <sup>V</sup> DATGH <sup>D</sup> QL <sup>A</sup> R <sup>L</sup><br>L <sup>E</sup> RR <sup>R</sup> PG <sup>S</sup> LS <sup>K</sup> V <sup>P</sup> GM <sup>S</sup> SLD <sup>V</sup> WT <sup>G</sup> ERQ <sup>V</sup> VE <sup>H</sup> TE <sup>G</sup> IFP <sup>G</sup> LYAAG <sup>M</sup> S <sup>V</sup> AEV <sup>V</sup> N <sup>L</sup> R <sup>R</sup> M <sup>G</sup> PF <sup>V</sup> FGG <sup>M</sup> IASA <sup>R</sup> LA <sup>E</sup> M <sup>L</sup> AER <sup>L</sup> AG <sup>K</sup> R <sup>M</sup> GLAT <sup>G</sup> VAR <sup>S</sup> G |

>Saccharicrinis fermentans DSM 9555 (2 Cys, 10 Met)  
**M**EQIVSVGIVDVF~~K~~KLKENLTVDVAIVGGGPGSMVAA~~Y~~YLARQGFKVSV~~E~~RKLAPGGGMWGGAMMFNEIVI~~Q~~KEALPILDELNISYKHYDKDYYT  
 LDSVHATSALIYHATQAGATFFNCTSVEDVVFLDNKVSGVVLNWAPVREKMHVDP~~L~~VIMAKAVIDGTGHDCDIARILERKNNIQLLTASGV~~E~~GER  
 SLSIDEAERTTIENTKEIYPGLYVSGMASNGVSGGFRMGP~~I~~F~~G~~GMILSGKKVANLIADNLNK

>Thermotoga maritima strain ATCC 43589 (2 Cys, 8 Met)  
**M**RDVLISRL~~V~~FEKLRNSLELDVAIVGAGPSGLTAAYELAKNGF~~R~~VAVFEERNTPGGGMWGGMMFNEIVL~~E~~KELENFLIKEVEIEYEVKEDHIV  
 VDSVHFASG~~L~~YRATKAGAIVFNNVSVEDAVQNGRVCGV~~V~~VNW~~G~~P~~T~~RLGLHVDPITVKASFVVDG~~T~~GHPANV~~S~~LLAKRGLVEMKTEFPMDADEA  
 EKFVVDNTGEIFPG~~L~~VSGMAVCAVHG~~G~~PRMGP~~I~~F~~G~~MLSGQKVARI~~V~~SERLR

>Thermovibrio ammonificans strain DSM 15698 (3 Cys, 10 Met)  
**M**QNLSSEV~~V~~ISEAI~~T~~AFMEKLKSH~~E~~TDVAIVGGGPG~~L~~VAG~~Y~~YLA~~K~~KG~~Y~~RAV~~I~~F~~R~~RLS~~I~~GGGMWAGAMFFNEIVVQEMGREILDEFGVNYREFKP  
 GY~~Y~~LA~~D~~AVEATTIASKAVKAGATV~~F~~N~~G~~V~~T~~A~~E~~D~~V~~V~~L~~K~~Q~~V~~N~~Q~~Y~~R~~V~~C~~G~~L~~V~~IN~~W~~T~~T~~V~~E~~L~~H~~LM~~V~~D~~P~~L~~V~~ITAKYV~~V~~DATGHDASV~~V~~STLQRKAGIKLNTE  
 TG~~V~~V~~G~~E~~K~~P~~L~~W~~A~~S~~V~~GE~~D~~DT~~V~~K~~N~~S~~K~~E~~V~~F~~P~~G~~I~~Y~~V~~S~~G~~MA~~N~~ATCG~~S~~H~~R~~M~~G~~P~~V~~F~~G~~GGMLMSGKKV~~A~~E~~E~~IA~~A~~KLQN~~N~~KEA

>Verrucomicrobia bacterium (1 Cys, 11 Met)  
**M**LN~~E~~VTISRAI~~D~~AYFKKLTRH~~L~~EV~~D~~V~~A~~I~~V~~GGGPG~~S~~LVAG~~H~~DLARAGKKVALFESK~~K~~LAIGGGIW~~G~~GGGMGF~~N~~EIVVQEAAREML~~V~~E~~F~~GLR~~A~~TEFEPGY  
 YT~~I~~DAV~~H~~VAA~~A~~LA~~A~~RA~~M~~EAGLTV~~F~~N~~L~~TS~~M~~ED~~V~~V~~I~~Q~~K~~DR~~V~~AGL~~V~~LN~~W~~T~~A~~IR~~H~~L~~K~~W~~H~~VDPLT~~I~~H~~S~~R~~F~~V~~L~~DATGHPAS~~V~~AETLVR~~K~~M~~N~~VR~~L~~DT~~T~~GG~~L~~VG  
 EK~~M~~AA~~E~~EDGERQTVENTREV~~V~~PG~~L~~F~~V~~SG~~M~~AA~~I~~TV~~C~~G~~G~~H~~R~~M~~G~~P~~V~~F~~G~~GG~~M~~LS~~G~~R~~K~~AA~~Q~~MLAEL~~G~~T

>Candidatus Omnitrophica bacterium 4484\_171 (8 Cys, 9 Met)  
**M**LEETIISKAI~~D~~SYHNK~~L~~SS~~I~~DV~~D~~AA~~I~~CGGGPG~~S~~LV~~C~~AASLAA~~G~~KKV~~V~~L~~F~~E~~K~~K~~L~~SLGGGMWGGGMF~~N~~EIVVQKKAKKILDEF~~S~~V~~R~~T~~K~~KY~~K~~EN~~Y~~  
 YLAD~~S~~SEC~~V~~C~~A~~LGYN~~A~~V~~H~~SG~~A~~V~~I~~ING~~V~~FA~~E~~D~~V~~W~~V~~K~~N~~R~~I~~C~~G~~L~~V~~IN~~W~~S~~A~~A~~S~~AN~~L~~W~~D~~PL~~T~~V~~R~~A~~K~~F~~V~~DATGHP~~S~~E~~V~~V~~K~~V~~E~~K~~S~~GV~~K~~I~~K~~T~~K~~G~~V~~LG~~E~~K~~S~~M~~W~~A~~H~~A~~E~~NT~~E~~IK~~N~~TRQ~~I~~AP~~G~~L~~F~~V~~T~~G~~M~~CAN~~A~~V~~G~~AP~~R~~M~~G~~P~~I~~F~~G~~ML~~S~~G~~K~~K~~C~~A~~I~~IL~~S~~RL

>Poribacter~~a~~ sp. WGA-A3 (5 Cys, 10 Met)  
**M**DNLQP~~A~~PLR~~R~~ERDV~~T~~RIAREFYKEFDQ~~L~~IES~~D~~V~~I~~IVGGGPG~~S~~LV~~C~~AHDLATQ~~G~~F~~R~~TL~~L~~IEQ~~S~~LA~~G~~GGFW~~S~~GGY~~L~~M~~N~~K~~A~~T~~I~~C~~E~~PA~~H~~SILE~~N~~MG~~V~~PC~~K~~P~~V~~KD~~C~~AG~~M~~R~~I~~V~~D~~P~~H~~AT~~A~~LA~~S~~YEAG~~A~~K~~V~~LN~~I~~TR~~V~~D~~L~~L~~H~~GE~~G~~V~~L~~E~~G~~V~~V~~V~~N~~TTAE~~M~~AG~~H~~D~~M~~I~~H~~VD~~P~~I~~A~~LES~~R~~V~~V~~V~~D~~ATGHD~~A~~V~~V~~V~~G~~LL~~N~~Q~~R~~LYAT~~V~~PG~~N~~GA~~M~~W~~V~~AR~~S~~EA~~M~~V~~V~~D~~N~~T~~R~~EV~~F~~PN~~C~~F~~V~~T~~G~~L~~A~~V~~A~~AV~~D~~G~~S~~PR~~C~~G~~P~~AF~~G~~ML~~S~~G~~R~~R~~A~~AD~~L~~V~~R~~H~~K~~L~~K~~G~~E~~

>Nitrospira defluvii (6 Cys, 12 Met)  
**M**EEL~~A~~RS~~R~~K~~A~~CT~~A~~VEGEY~~R~~M~~G~~K~~P~~K~~P~~A~~L~~R~~R~~ER~~D~~IT~~R~~Q~~I~~ARE~~Y~~Y~~K~~E~~F~~D~~Q~~L~~I~~E~~S~~D~~V~~I~~I~~VG~~A~~G~~P~~GL~~I~~C~~A~~H~~D~~L~~G~~R~~M~~GI~~K~~T~~L~~IVE~~Q~~SL~~A~~LG~~G~~GF~~W~~SG~~G~~Y~~L~~M~~N~~K~~A~~T~~I~~C~~A~~PA~~H~~K~~I~~KE~~V~~G~~V~~P~~C~~Q~~I~~KE~~C~~PG~~M~~Y~~V~~D~~P~~PHAT~~G~~ALIA~~A~~AYNAG~~A~~K~~M~~LN~~L~~TR~~V~~D~~L~~L~~R~~REG~~V~~LEG~~V~~V~~N~~TTAE~~M~~AG~~H~~D~~M~~I~~H~~VD~~P~~I~~A~~LES~~K~~IVV~~D~~ATGHD~~A~~V~~V~~V~~N~~LL~~H~~K~~R~~GL~~Y~~Q~~Q~~PG~~N~~GA~~M~~W~~V~~SR~~S~~EE~~E~~V~~M~~D~~R~~T~~G~~E~~V~~S~~P~~N~~C~~F~~V~~I~~G~~L~~A~~V~~A~~AV~~F~~G~~T~~PR~~M~~G~~P~~AF~~G~~ML~~S~~G~~R~~Y~~G~~A~~E~~L~~R~~D~~K~~L~~K~~N~~R~~

>Desulfurococcus amylolyticus strain DSM 18924 (1 Cys, 5 Met)  
**M**SLES~~H~~IT~~R~~VI~~W~~E~~A~~SRD~~W~~VEL~~S~~~~C~~DIV~~V~~G~~A~~G~~P~~GL~~T~~A~~K~~Y~~L~~A~~E~~KL~~G~~K~~T~~L~~V~~L~~E~~R~~R~~L~~S~~F~~G~~GGG~~G~~GGG~~G~~ML~~H~~K~~T~~V~~V~~D~~E~~R~~G~~L~~G~~I~~L~~R~~D~~F~~N~~I~~R~~Y~~K~~P~~S~~I~~K~~G~~L~~LY~~V~~V~~D~~T~~A~~E~~L~~T~~A~~LA~~G~~AL~~D~~A~~G~~K~~I~~IP~~G~~I~~S~~ED~~V~~V~~I~~R~~Y~~N~~P~~FR~~V~~Q~~G~~V~~V~~V~~E~~SA~~V~~QL~~S~~GT~~C~~VD~~P~~L~~F~~I~~S~~E~~K~~A~~V~~IDA~~T~~G~~H~~D~~A~~E~~V~~L~~R~~I~~E~~K~~K~~N~~P~~E~~S~~K~~V~~K~~I~~P~~G~~E~~K~~KS~~A~~Y~~E~~K~~A~~D~~V~~D~~V~~E~~Y~~T~~G~~R~~V~~I~~P~~G~~L~~Y~~A~~T~~G~~MA~~V~~AA~~V~~R~~G~~L~~N~~R~~M~~G~~P~~I~~F~~G~~M~~IL~~S~~G~~R~~K~~V~~A~~E~~A~~V~~R~~D~~LES~~A~~PK

>Candidatus Aenigmarchaeota archaeon (2 Cys, 10 Met)  
**M**GBIIFSKV~~S~~E~~K~~V~~T~~SA~~I~~V~~S~~G~~F~~I~~K~~E~~F~~K~~I~~I~~E~~S~~D~~V~~I~~IVGGGPG~~S~~LM~~A~~G~~K~~E~~L~~SS~~K~~G~~K~~V~~V~~II~~E~~R~~N~~Y~~L~~GGG~~F~~WT~~G~~GY~~L~~M~~N~~K~~I~~T~~V~~R~~H~~P~~G~~E~~E~~I~~L~~K~~D~~L~~G~~I~~P~~FE~~E~~FG~~E~~GL~~Y~~LA~~D~~G~~P~~H~~A~~CS~~K~~L~~I~~A~~T~~TDAGV~~K~~IL~~N~~TT~~I~~LED~~V~~V~~L~~K~~E~~G~~A~~V~~G~~V~~I~~N~~W~~T~~P~~I~~E~~T~~P~~R~~E~~I~~A~~VD~~P~~I~~A~~LES~~K~~V~~V~~VI~~D~~ATGHD~~A~~V~~V~~V~~N~~K~~I~~E~~E~~RG~~I~~L~~K~~

>Pyrolobus fumarii strain DSM 11204 (0 Cys, 10 Met)  
**M**VI~~P~~G~~H~~M~~T~~TR~~R~~T~~A~~MP~~G~~L~~D~~AI~~I~~TR~~V~~I~~I~~E~~A~~SK~~E~~L~~V~~Y~~E~~A~~S~~D~~V~~V~~I~~V~~G~~G~~A~~G~~P~~GL~~T~~A~~A~~F~~Y~~LA~~K~~R~~G~~F~~R~~V~~L~~V~~L~~E~~R~~R~~L~~S~~V~~GG~~G~~GG~~G~~ML~~F~~H~~K~~V~~L~~V~~Q~~E~~E~~AL~~P~~V~~L~~ND~~M~~G~~R~~IV~~H~~P~~T~~SV~~K~~GI~~Y~~SL~~D~~S~~V~~AL~~I~~IT~~G~~L~~A~~S~~A~~A~~V~~N~~A~~G~~A~~K~~I~~IL~~G~~LE~~A~~V~~D~~V~~L~~V~~R~~KE~~E~~HR~~R~~V~~A~~G~~V~~MA~~L~~W~~S~~AV~~G~~I~~A~~N~~L~~H~~V~~D~~P~~L~~F~~E~~A~~K~~V~~V~~D~~ATG~~H~~D~~A~~E~~V~~L~~R~~IA~~H~~Q~~K~~L~~R~~GE~~A~~P~~V~~PG~~D~~G~~P~~WA~~A~~E~~E~~GE~~K~~L~~V~~V~~K~~AT~~G~~E~~L~~P~~I~~G~~L~~Y~~V~~A~~G~~MA~~T~~A~~T~~V~~K~~G~~Y~~Y~~R~~MG~~P~~I~~F~~G~~G~~ML~~S~~G~~K~~K~~V~~AD~~L~~ITE~~K~~L~~R~~G~~K~~

>Metallosphaera sedula strain ATCC 51363 (0 Cys, 8 Met)  
**M**NI~~K~~Q~~V~~D~~E~~I~~K~~I~~T~~RY~~L~~K~~A~~T~~F~~ED~~W~~DFS~~V~~N~~D~~V~~V~~I~~V~~G~~A~~G~~P~~GL~~A~~AA~~Y~~Y~~S~~A~~K~~AG~~L~~K~~T~~T~~V~~F~~E~~R~~R~~L~~S~~F~~G~~GG~~G~~GG~~G~~ML~~F~~H~~K~~V~~I~~V~~E~~S~~P~~A~~D~~E~~I~~L~~R~~I~~E~~G~~V~~K~~V~~LF~~Q~~K~~F~~EE~~G~~V~~V~~V~~D~~S~~S~~EF~~N~~AA~~K~~L~~A~~A~~T~~TI~~D~~A~~G~~K~~I~~II~~H~~G~~V~~T~~V~~D~~V~~I~~F~~R~~E~~N~~P~~L~~R~~V~~G~~V~~A~~E~~V~~WT~~T~~Q~~A~~SL~~H~~I~~V~~D~~P~~L~~F~~I~~S~~A~~K~~V~~V~~D~~A~~T~~G~~H~~D~~A~~E~~V~~I~~S~~V~~A~~R~~K~~I~~P~~E~~L~~G~~I~~V~~P~~G~~E~~K~~S~~A~~Y~~E~~I~~A~~Q~~L~~T~~V~~E~~Q~~S~~G~~E~~V~~A~~P~~G~~L~~Y~~A~~G~~M~~AV~~T~~E~~I~~K~~A~~IP~~R~~G~~P~~I~~F~~G~~M~~LL~~S~~G~~K~~K~~V~~A~~E~~D~~I~~I~~K~~N~~L~~Q~~A~~S~~T~~L~~K~~S~~V~~Q~~K~~E

>Caldivirga maquilingensis strain ATCC 700844 (0 Cys, 10 Met)  
**M**AGISIRES~~A~~IT~~R~~AI~~V~~N~~S~~A~~K~~L~~L~~SE~~Y~~SS~~V~~D~~V~~AI~~V~~G~~A~~G~~P~~SG~~M~~TA~~A~~YY~~L~~A~~K~~G~~L~~K~~T~~L~~V~~L~~E~~R~~R~~F~~S~~F~~G~~GG~~G~~GG~~G~~AA~~SH~~L~~P~~SI~~I~~V~~E~~H~~P~~V~~S~~E~~I~~L~~S~~K~~D~~F~~G~~I~~K~~IM~~D~~MG~~D~~G~~L~~F~~T~~V~~D~~PA~~E~~MI~~A~~K~~L~~A~~V~~K~~A~~I~~D~~A~~G~~K~~F~~L~~G~~V~~H~~V~~D~~D~~V~~I~~F~~R~~E~~N~~P~~L~~R~~V~~G~~V~~A~~E~~V~~WT~~T~~Q~~A~~AG~~V~~H~~P~~F~~I~~S~~N~~AV~~V~~D~~A~~T~~G~~H~~D~~A~~E~~V~~I~~S~~V~~A~~R~~K~~I~~P~~E~~L~~G~~I~~V~~V~~R~~GE~~K~~S~~A~~Y~~V~~G~~V~~A~~E~~D~~L~~V~~V~~K~~T~~G~~V~~ID~~G~~L~~Y~~V~~T~~G~~M~~MA~~V~~A~~A~~V~~H~~G~~L~~R~~M~~G~~P~~I~~F~~G~~M~~LL~~S~~G~~K~~K~~V~~A~~E~~D~~I~~I~~E~~DL~~K~~GN~~H~~

>Methanofollis liminatans DSM 4140 (5 Cys, 12 Met)  
**M**ELDEV~~T~~IS~~R~~AIL~~A~~T~~Q~~M~~E~~V~~Y~~LD~~D~~V~~V~~V~~V~~GGGPG~~S~~IT~~C~~A~~L~~LA~~E~~K~~G~~V~~K~~V~~G~~L~~F~~E~~K~~K~~L~~S~~I~~GGG~~M~~W~~G~~GG~~M~~F~~P~~R~~I~~V~~V~~Q~~A~~E~~K~~R~~I~~L~~D~~R~~F~~G~~I~~A~~S~~K~~E~~F~~E~~PG~~Y~~H~~V~~A~~K~~S~~V~~E~~A~~V~~S~~K~~L~~AA~~A~~ACT~~A~~G~~E~~E~~F~~N~~L~~I~~A~~V~~E~~D~~V~~V~~I~~K~~G~~D~~G~~R~~L~~AG~~L~~V~~G~~V~~N~~SP~~V~~E~~M~~AG~~L~~I~~T~~D~~P~~L~~T~~I~~R~~CK~~A~~V~~V~~D~~S~~G~~H~~D~~A~~T~~I~~A~~H~~W~~V~~A~~K~~K~~G~~G~~D~~L~~P~~I~~R~~G~~E~~G~~F~~M~~W~~AD~~R~~A~~E~~G~~N~~I~~L~~E~~H~~T~~R~~V~~F~~P~~G~~L~~F~~V~~C~~MA~~N~~A~~A~~N~~A~~V~~A~~G~~E~~CR~~M~~G~~P~~I~~F~~G~~G~~ML~~S~~G~~K~~K~~V~~ERA~~D~~LA~~A~~A~~A~~V~~L~~H~~P~~

>Sulfurisphaera tokodaii strain DSM 16993 (0 Cys, 10 Met)  
**M**DSNSIKV~~K~~Q~~V~~D~~E~~V~~K~~I~~S~~K~~Y~~I~~L~~K~~T~~F~~Q~~D~~W~~E~~I~~V~~S~~D~~V~~V~~V~~G~~A~~G~~P~~SG~~M~~TA~~A~~YY~~L~~A~~K~~G~~L~~K~~T~~V~~V~~F~~E~~R~~R~~L~~S~~F~~G~~GG~~G~~GG~~G~~ML~~F~~H~~K~~V~~I~~V~~E~~S~~P~~A~~D~~E~~I~~L~~K~~EM~~K~~I~~K~~N~~V~~E~~E~~G~~V~~Y~~I~~V~~D~~S~~A~~E~~F~~MA~~K~~L~~A~~A~~S~~I~~D~~A~~G~~K~~I~~II~~H~~G~~V~~T~~V~~D~~V~~I~~F~~R~~E~~N~~P~~L~~R~~V~~G~~V~~A~~E~~V~~WT~~T~~Q~~A~~AG~~V~~H~~P~~F~~I~~S~~N~~AV~~V~~D~~A~~T~~G~~H~~D~~A~~E~~V~~I~~S~~V~~A~~R~~K~~I~~P~~E~~L~~G~~I~~V~~LN~~V~~I~~P~~GE~~K~~S~~A~~Y~~E~~I~~A~~E~~E~~L~~T~~V~~E~~N~~T~~G~~M~~V~~A~~P~~G~~LY~~A~~G~~M~~AV~~T~~E~~V~~K~~G~~L~~P~~R~~G~~I~~F~~G~~M~~VL~~S~~G~~K~~R~~V~~A~~E~~I~~I~~I~~K~~D~~L~~R~~Y~~

>Deltaproteobacteria bacterium HGW-Deltaproteobacteria-1  
MVLDEIVISKAIIERFLEKLLQATDVDVAIVGGGPGSLVAAAYYLASAGKKVALFERKLSLGGGMWGGMMFNEIVVQDEAREILDVFDIYREYQQG  
YYTADAVLAVTSICSAARAGASIFNCVSVEDVMIREGRVTGLVINWSPVEMAGLHVDPDTIAAGSVIDTTGHATEVLKIERKADMQLATPSGKLV  
GERSMWAekaerltmdntrqicpgvvyagmsanaafggprmgpifggmllsgrkvaeillass

>Thermogladius calderae (strain DSM 22663 / VKM B-2946 / 1633)  
MELESIITRLVVEESARELVELSESVDLVVGAGPSGLTAALKYLADKHVKVVLEKRLSYGGGIGGGGSLFHKVVVDERALPVLGDFKVRYKAAGVAG  
YYVVDSAELMSKLAAGALDSGAKIILGAEVEDLVRDNPLRVRVGVFMFKWSAITAAGLHVDPFLFALSRAVVDATGHEAVLVSILSRKNRVAGVAVPGE  
RSGFAERAERDVVEYTGRMPGLYVAGMSVAAVHGLHRMGPIFTGMLLSGRKVAEAIARDLGVPQ

>Acetomicrobium thermoterrenum DSM 13490  
MKLDELVITKAIVEGYFKLMMCLEDVAIVGGGPGSLVAAALELAKAGKKVALYERKLSVGGGMWGGMLFNEIVIQHEAKEILEGVGVNRPYEVE  
GYYTADSVEAVSTLTSKAVKAGATIFNALSVEDVVVDEERINGLTVNNWTAVEMAGLHVDPFLSIHCKYIDATGHDTEVVRRVARKMPGRLFTATGN  
IEEGKFMSPDRAEKLTIVNTREVFPGLYVAGMAANATFGGPRMGPIFGGMLLSGVKAAREILSK1

>Acidianus hospitalis (strain W1)  
MQSIRIKQVNEVKISKYILKYTFEDWNNLVESDVVIVGAGPSGMATAAYYLAKAGLKTIVFERRLSFGGGIGGGAMNFHKVIETPADEIIKELKIRY  
IEPEEGIFIIDSAEFMAKLATAAIDAGAKIIHGVTDDVIFRENPLRAGVAVEWTSTQMSGHLHVDPFLFISAKAVVDACTHDAEIISVASRKVPELG  
IAVPGEKSAYSEIAEELVVENTGKAVPGLYATGMACVEKSLPRMGPIFGAMILSGKKVAEEIIKDLRNS

>Acidithiobacillales bacterium SM23\_46  
MCCQSLEAWRSEPEERKRNRADVVVVGAGPSGMATAIHLARERHRVILLEKRLSPGGGIWGGGMAMSEAIIVQDDALPWLDLGVRHKPSRGGLHSADA  
VELAAALCLKTVQSGTFLFNLLTVEDVCIHQDRVTGVVNRSMIAGALPVDFIAFRNAVIDATGHEAVVVEAVHKRGLLAHPAVAKPLGEGPMADA  
SGEAFVVENVKEVYPGLWICGMSVCATLGPRMGPIFGMLLSGQRVAALVSSALTEFAQKDRESRK

>Aciduliprofundum boonei (strain DSM 19572 / T469)  
MLDEVEITKLIVENYMKDLMEYADLVAIVGAGPSGLTAAYYLATAKKVAIFDRRLSFIGGMWGGMMFNKIVVQEDAKHILDDFSINYERFGDYY  
VADSVHSVTSLAYHATKEGAKIFNLIGAEDVIICKNNRSGLVINWSVIGELPIDPLSIYAKYVIDATGHESEVIKTLVRKNNIKLNTPTGSIEGEHS  
MDADTAESVIVDNVKEVYPGLFVTGMAANAVFGSPRMGPIFGGMILSGKKVADEIIIRRLS

>Actinobacteria bacterium HGW-Actinobacteria-3  
MPLSEIEVTRGILEGFSRDFLSSLQSDVAIAGAGPSGMVCAYYLAREGLKVSFERNLHVGGGMWGGMLFPRIIIQEAAIREVEFGVRLKPFKEG  
YFVGDSVETVKVTAACIDAGVRVWVGVSVEDVLIREEENRLAGVVLNWRABELANLHVDPЛАVEAKVVDATGHEAVVRTVARKIPGCRINTDTGG  
VIGEMPMWAQVGEELIVGNTREYANLLVTGMAANAVYGAJPRMGAIFFGMFLSGYKCAHLAADIVRKA

>Alistipes inops  
MIETKVSQGIVSTYFDKLQKNLELDVAIVGGGPGSLVAAAYYLAKAGLRLAQGLKLTVVFERRLSVGGGMWGGAMMFNYLFQEEARPIFETMGVRYREY  
VMDVESTSALLYHAVHAGATVFCNCYSEDVYVKENRSGVVNVTPVLRGLVDPNILARVVIDGTGDSEIAATVARKNGARLNTEGTVVGE  
RSLDVTAGEDEVVKGTKEIYPGLYVCGMAASAVSGTPRMGPIFGGMILMSGKKVADEIIARLKK

>Ammonifex degensii (strain DSM 10501 / KC4)  
MAGGAIDERLVSRAIIQTYSEELLQLTDFDVAVVAGPSGLTAAYYLAKAGLRLAQGLKLTVVFERRLSVGGGMWGGAMMFNYLFQEEARPIFETMGVRYREY  
QPGYYVAHSVEAAFTLAACRAGARIMNLITDVLRDNRVAGLVLNWTAVDMAGMHIDPLAVHCRYVVDATGHDAEVVRILTQKNQVTVKVPGG  
HVQGEKSMWSERGEKQTLDHSGEVFPGLYVAGMAANAVAGGYRMRGPIFGGMVLSGKKVAELILEAHRREKSQTL

>Ancylomarina sp. 16SWW S1-10-2  
MEQIVSAGIVDSYFKKLKENLSVDVAIVGGGPGSLVASYYLAKKGFKVALYETKLAPGGGMWGGAMMFNEIVVQKDALHILDELNSYTNYQGDYYT  
LDSVHATSALIYHATQAGVKIFNCSSIEDVVFQNNKVCVVLNWSPRREGLHVDPFLVIMAKAVVDGTGHECDIVSTLERKNGVKLNTKTGKVMGEC  
SLSIDEAERTTVENTKEVYPGLYVSGMASNGVSGGFRMGPIFGGMILSGEKGKLAGLIAENLSK

>ANME-2 cluster archaeon HR1  
MELDEITITRAIILEDFTSDFLQSIDTDVALVGGGPANLIAARTLARAGVKTFLFERKLEVGGGMWGGMMMPRIIVQEEARHILDDLGVRYRKYEEG  
YYVADSIECTGKLIYEASSGASIYNLISVEDVMIREGRDAVTGLVINRTIVDMQKLHVDPITIRAKVVIDGTGDSEICTTLSRKIPGALHVAGEKP  
MWADVAERIILDNTKEVYPGLIVTGMAANAVAGAPRMGPIFGGMILSGEKAQIAIAKGL

>Archaeoglobales archaeon ex4484\_92  
MEARIKSAIIIEVAKDWSNISQVDVIVGAGPSGLTAGKYLAEKGLTLILERRLCFGGGIGGGMFLHKIVIEKFAREILDFFDVRYYEHDNLLVA  
DVAEFLMAKLAVGCVNAGTKIIGHGSVCDVIFRLEPIRITGCVCIQWSAELSGLHVDPMFIESKAVLDATGHDAEVVSIAASKVPLDLNVGTGEKSAYA  
ELGEKLVVEKTGKVVEGLYATGMACSVFNLPRMGPIFGGMLQSGKKAAEIIYNDLK

>Archaeoglobus fulgidus  
MEAETKAIVETASEEWVEYAESDVIVVAGPSGLTAARYLAEKGLTLVLERRLSFGGGIGGGMFLHKVVVEREAKDILDDFGIRYTEHRNFLVA  
DSAEMAKLAAKIAIDAGAKIIGHGSVVEDVIFRDDPLGVRGVCIQWSAELSGLHVDPFLRSRAVVDATGHDAEVISAAKIPLEVSVVGERSAYS  
EVAEREIVEKTKIVKGLYAAAGMAAAVHNLP RMGPIFGGMILSGKKVAEIVADELKL

>Armatimonadetes bacterium  
MSFEKDSFKWDELTIVTRGIVETFMADFLSIDLDVAIVVAGPSGLTAARILAGQGHRVGIFERNLHIGGGIWWGGMLFPRIIEEEAAMPLMEAAGVK  
LRPWKDGTVIADAVESATKMTAAIIDAGARIFVGIEADVVVDDSDRVCVGVINWGAVTAAKLHVDPFLAVHSKVLTESTGHPCEVGDVLLRKIPGAR  
LDTGETCPGEASMNARAGEAALIANTREIYPGVVVAGMAANAVSRSPRMDAIFGGMILSGQKAAEISAQIIADLG

>bacterium (Candidatus Ratteibacteria) CG23\_combo\_of\_CG06-09\_8\_20\_14\_all\_48\_7  
MLDDVVISRAIETYFQFDLLNYLENLDVIAAGAGPAGLTAAYFLAKKGRKVAFQERQLRVGGMPGGGMFNKIVIQEETKKILDEFGIRYQKYQNGY  
YVADSLLETSLLTSKAIQAGAKIFNLIAVEDLSQEGVRNVGLNWSAVTAGLHVDPITIRAKAVVDAKGAVLCRLLVDKGVTLRPTGKVA  
EGPMAKEGEEMIANTGEVFPGFLTAVGTMVNAVCGGPRMGPIFGGMILSGERLSSLIP

>bacterium (Candidatus Stahlbacteria) CG23\_combo\_of\_CG06-09\_8\_20\_14\_all\_40\_9  
MMLEDTIISRAIIETYKEFKVNLKSVAIAGGGPSGLIAGYVYKKKKPDLKVLFERKLISIGGMWGGMMMNEIVVQEEGKKILDEFGVKSKKFE  
NGYYTADSIEVTSLALANTVKAGVTILNAISVEDTIIDNIAKGLVINWTSALDIGHVDPDLRADHIIDATGHPCEIAHLLIEKKGKLFRTGKI  
IPEGAMYADKGVERVIENTKEFPNVWACGMAANAVFGGERMPGIPEGMLLSKGKIVAEKILOKN

>bacterium\_42\_11  
MKDILISKAILESFNNKLRDSLELDVAIVGAGPSGLVASYELAKKKVKIAIFEERNTPGGGIWGGGIMFNEVVLEKELEDFLKELDIKYKVVEDYIV  
VDSTHSAFALIYHTTWGTRIFNSISVEDIAMQNRRVCGVVINWGPVKKLGLHVDPITIKASYVVDGTGH PANVVSSLVKRGLLEKKTEFPMNAEAA  
EKFWVVEKTGEVFGPLJLVSGMAVCFVYGGPRMGPTFGGMVLJSGRTRIAETTTERVNKR

>Bacteroidales bacterium 6E  
MEQIVSSGIIIDSYFKKIKESLSVDVAIVGGGPGLVAAAYLAKKGLKVAMFERKLAPGGGMGGAMMFNEIVVQKGALQILDEFKIDYTHYEGDYYT  
LDSSQATSTSLIHYHAGKAGARIFCNCTSVEDVVVFHNKNVCKGVINVNAPWVHRERLHVDPVLVIMAVIDGTGHDCDIARILERKNNIQLNTVSGKVQGER  
SISDEAFTPTVYENKEFPLCYSCMAANCYSCCERMCPTECMJLISCEKVIALIMNOTNK

>Bacteroides cellullosilyticus  
MIETKVSKGIISTYFEKLERNLDVAIVGGGPGSIVAAAYLAKAGLKVAFQFDRKLAPGGGMWGGMAMMFNQIVIQEEAIDIVKEFNINHEKYEDGLY  
DMSVSESTSVALLYHAVHAGATVFNCVSYVEDVCFVNKNNTSPGVVVNWTPVLRGEHMVDPLNLLAKIVIDGTGHDSEIAATVARKNGSRLATEGGVIGE  
DGIKVEVCGEEVWVNGTIVLWGLVNUCCGIVCNMCTSDPMLIEGCGMLMCGKXWILRDIILKIIKX

>Bacteroides sp. 3\_1\_19  
MEQIVSSGIIDSYFEKLKSNLNSVDIAVGGGPGSGIVAAYFLAKAGKKVALFDRKLAPGGGMWGGAMMFNDIVVQEEAMPIIKELGVSYKEGANGYI  
MDSVHTTSALIYQATKAGATIFNCYSVEDVVVFHNDAVAGVVVNWAPVIREGMHVDPPLTIMAKAVLEGTGHDCIEIARVARKNDIQLNTPTGGVIGER

>Bacteroides stercorisoris  
MIETKVSKGIISTYFEKLERNLDLDAIVGGGPGSGIVAAYYLAKAGLKVAFDRKLAPGGGMWGGAMMFNQIVIQEEAIDIVKEFNINHEKYEDGLY  
VMDSVESTSALLYQAVHAGATIFNCYSVEDVIFKNNTVNGVVNWTPVLREGMHVDPLNILAKVVVDTGHDSEIAATVARKNGIRLATETGGVIGE

>Bacteroidetes bacterium ADurb.Bin035  
MEQIVSNGIIDSYFNKLQYLSVDVAIVGGGPSGLVASYYLAKNNYKVAIYERKLAPGGGMWGGAMMFNEIVVQKAALHILDELSIEYREFEHDFYV  
IDSVHAASALIYNASKAGVKIFNFNTSVEDVVFLDNKVCVGIVLNWAPVAREHLEIDPLVIMAKVVVDSTGHDCDVAHTLERKNNIKLNTETGKVIGER

>Brockia lithotrophica  
MFDERVITRAIVETYLEEFRSIVDLDVAVVGSGPSGLVAARELARRGYRVAVFKEKRLSVGGGLWGGGMLMNRIVVQEAAPILEEFGVRTREYAPGV  
YVAHSVETVAALVFGALQAGAYVFNGVAMEDVVVRDGRVAGLVLNWGAHAAGLHVDPALIHARAVLDATGHDAEVVRKLAAKNGVSLVSGERSMW

```
>Caldimicrobium thiodismutans
MELQINQAIIREGMRDLEDSDVVLIAAGAGPSGLTAALKYLAERGFKVLITYERRLSFGGGIGGGGNMIPKIVVQTEALPIVEDFKIRAKKVENGFF
IDPAELIAKLATGALDAGAKIFLGVNVDDIVRDAAPRUVVGVLWHWTQIQLSGLHVDPYLTHCKALVDATGHDAELIAIAGRKNPELGIEAGEKS
```

>candidate division MSB11 archaeon SCGC-AAA259E19  
MLDEEVITKAIVEEYMTFLENTDVEAALGGAGPANLVAAKKLAEKGVKTAVYEELNVGGGMWGGGMYPRIVVQEAEKRILKEFGINFSEYEKG  
YVASSIESVARLASEFAVKAGAEIFNLTKVEDVMVREDDHIAGVVLNWSAVEKANLHVDPPLVKADVVIDGTGHDAEICRVTORKIPIDADLOVRGEKP

>candidate division TA06 bacterium 78  
MIEETIISKAIVESYSLKDVSIESDVIYIAGAGPSSG\_78  
FVADSLLETACILTEKALKIAGLKTFNIVSVEDVMTREGCVTGWVNIWSAVERMARIHDPTSEFSKVVIDATGHPSEIVHIVKEKKSGGKLITPSGRIG

>candidate division WOR\_3 bacterium SM1\_77  
MIEETVIRSRAIIESYLNLDLSDIOSDVIIGGAGPGLCASYYLAKKGHKVVLFERALKLGGMGGGIMFNRIVVOETARRILDEFairyEEYKPGY

>candidate division Zixibacteria bacterium 4484 93

LYIADAVESVGALIYRSTQAGLRLIFNLTIVEDLLVKDGKVEGLVINWSPVEMASLHIDPLTVQSRYSVDATGHPLEVVRTLCKKSGVQLFTETGDVL  
GERSMCADEGKFVLEKTGEVAPNMFVAGMAACAAFGGPRMGPFIGGMILSGKKVADIITKRLKEKK

>Candidatus Acetothermia bacterium  
MKIDDVLVSRLLIEEYMAFDLDCLHIDVAIVGAGPAGLTAAYYLAKAGAKVAVYERKLAIGGGMWGGGAMFSRIVIQEEAKQILDQFKITSIAKEGDYYVADAIEAITLLAAGAIQAGAKVFNLIHIEDLLLNRDVRVEGLVLQWSPVEMGGHLHVDPITIGAREVIDATGHCEVVKKLLHKGGVKIDTETGGMLGERPMWAEKGEKMTVQYTKQIYPGLYVAGMAVGAVFGTPRMGPIFGGMLLSGKKAQIIAQTLS

>Candidatus Acidianus copahuensis  
MKVKQVDEGKISKYILKFTFEDWENIIDSVDIIVGAGPSGMAAYYLAKAGLKTFLFERRLSFGGGIGGGAMLFHKIIIESPADEILKELGIRLVKAEDVYAVDTAEFMAKLASSAIDAGAKFIHGTVDDVIFREEPLKVAGVAVEWTATQMSLHVDPIFISAKAVVADATGHDAEVISASRKIPELEISIPGEKSAYSEVAEQQVVDGTGKAPGLYAAAGMAVCEIKGLPRMGPIFGAMVLSGKKVADEIINDIRKS

>Candidatus Altarchaeum sp. CG2\_30\_32\_3053  
MIEETKITELIVRNAVDDFLNNLDVvvVAGGGPAGLTARYLAKAKKKVVLFERKLSIGGGMWGGMMFPRVVLQKGGKEILEECDVCKKSGGLWVADSIECVTKMTAKAIDEGVKIFNLVSIEDVIIRNNEDNKNKNRKNKTCIGVVIINWTAQVOMANLHVDPLSVSKFVVADATGHEASISHLVVKVGNLNTKTDVGLERSWAEKGESDIMGNTGEVYVPGFVSGMAANAVYGSERMGAIFGGMLMSGKKVSELILKKI

>Candidatus Aminicenantes bacterium 4484\_214  
MIDEIVISRAITEAYLKEFLDCLESVDIISGAGPAGLCAALNLAQEYKVVFERTLRPGGGVPGGMMFNKIVIQEEARPLLEELDVTLKPYQENYYVVALELLGALLVKAIKQGVYLFNCISVEDVLIYDKKVGVVINWSAAQAAGLHVDPARTKFVVDATGHAEVAEIVSRKGCKLFTSTGKVTGEKPMWAEEGEKILLNTKEVYVPHLYVCGMAANAVFGGPRMGPIFGGMLLSGQKVAQLIAARLKT

>Candidatus Aramenus sulfurataquae  
MQSIRIKQVDEVKISRILYKTFEDWYSLVSDSDVVVAGGPSGLATAFYTAKAGLKTFFERRLSFGGGIGGGAMNFHKIVIESPADELLREWKVKLVEAEEGVFIVDAAEFMAKLGAAIAIDAGAKVIGHNIVDDVIFRDKPLRVAGVAVEWTSTQMSGHLHVDPFLVSAKAVVADATGHDAEILSVASRKIPELGIVIPGEKSAYSEVAELVNNAGKVAEGLYTTGMAVCEVKSLPRMGPIFGAMVLSGKKVAEDIINDLRNS

>Candidatus Aureabacteria bacterium SURF\_26  
MPLDDLIVSRAIIDYEHTTLVDALNMDSVAIVGGGPAGLIVAGYLYLAKQGYKVSLFERKLSIGGGMWGGGIMFNKIVLQEDALKVLNEFNRVKKYRDNYYVADSVEVGALIYHATQAGLQIFNCMSIEDVKITADAVCGLVNNWTSVELTNMHVDPTFGAKFVIDATGHSCDMANIILKRIGKVLFTPTGDMGEGSMAELAKVVAENSREIYKNLYVTGMAANAIWGSKRMGPIFGGMLLSGKKVADDISARLAQEAVN

>Candidatus Bathyarchaeota archaeon B26-2  
MRIKEVDEAVVTKAILEGSSLKYLHELTEVDVAVVAGAGPAGLTASRYLAKAGLKTFFERRLSFGGGIGGGMQLPMLVVQSPADEFILREVCNLTTYREGVYLANSSLEMALKLAGVSAKAGAHILGTVDDILYRSEENRTRIVGVVVQWSSVIISGLHVDPFLAKAGAVVDCGHDAEVLVASRKIPELNMIQGEKAMWVSESERLIVEKTGEVSPGLYVAGMAVATLNQTPRMGPIFGGMLLSGKKVAEIIERYKQNRTQ

>Candidatus Desantisbacteria bacterium CG2\_30\_40\_21  
MQLDDVVISRAIIIESWNKDLDSLIDVAIVGGGPAGLTCGYLSKGLKVVLFERNLISIGGGMWGGMMYNKCVFQQESLPILNEFGIRTQEYQDGYFTDSLETVTLCGALKAGLKIFNLIGVEDVMIRQEGVTGLVNWTAVTMAKLHDPLTIRAKAIVDSTGHAAEVAGIIVRKIGKKLLTETGEMLGEKPMWAEVGERTIAENTKEIYPGVVFAGMSANAVFGGPRMGPIFGGMLLSGKQAAEIIAARL

>Candidatus Desulfuridis sp.  
MKIDETVVSRAIIERYTQKLLSCLDVDEIVGGGPAGLTAAYHLAKHGKVTTLYERKLSVGGGMWGGAIMMNEIVFQEQARPLFEFGIRINPYSDGYYTASSVECAALTQACQAGANIINLMTVEDVVLHEDRVSGLVLNWTAVDIAGLHVDPATRSKYVIDCTGHMEVANILSKAGVKLVTPSGEPVGEKPMWADVGENQLIGHTIEVYVPGFVAGMAANAVNGGYRGMGAIFGGMVLSGRRAGELILKRLQS

>Candidatus Fermentibacteria bacterium  
MENIVTSAAEDYHRKIQESTVTAHVAIAGGGPSGLVAAEALASRGISVNLYEKNLTPGGGMWGGAMLFNSILIEEEFAETAELGMKLRKYRDSVLLADSVQATAALISRACSAGVRMFNGMAVEDVTVDYDRVNGVVNWAPVMKLGMMVDPLMTCAGAVLDATGHPAEIVTRFAAKNNTIEITVPGEASLNVEMGEKHTVEHTGMVHPGLFVSGMSACATAGGYRGMGPVFVGGMLKSLKAAEITEYLNKSQ

>Candidatus Korarchaeota archaeon  
MFVVEESVISSAIIERGSKFLVDLVSVDIIVGAGPSGLVAGRYIAKGLKTCIERRLSFGGGIGGGMLFPRIIVVQEPAQEILEEVGVKLEPYSKGVWIADVAETIKAAGAIDSGARILLGANVEDLIVRNSRVCVVVQSAVTSAGLHVDPFLAFESRAVIDCTGHNAEVVAAIAARKNPELGIRVLGEHSMDAVRAEKEVVELTGEVLPGLWVAGMAAAAVRGGPRMGPIFGGMLLSGKKVAELVSRELEV

>Candidatus Nitrospira inopinata  
MHKPKPAPLRLERDVTRHIAREYYKEFDQRIESDVIIVGAGPSGLLCAHDLAAMGFRTLIVEQSLALGGFWHGGYLMNKATICEPANEILEEVGVPCKRIACDCGMYMVDPPHATGALVAAAYRAGAKILNLTRVVDLILRQDGLLEGIVVVNNNTAEMAGHDVIIHVDPIALESKIVVADATGHDAVVLVELLHKRNLYKPVPGNGAMWVSRSEEEVMDRTGEVYVNCFVIGLAVAAYVTPRMGPAGFSMILLSGRYGAQLIKKKLKQE

>Candidatus Nitrospira nitrificans  
MAKPRPAPLRLERDITRHIAREYYKEFDQRIESDVIIVGAGPSGLICAHDLAAMGFRTLIEQSLALGGFWHGGYLMNKATICEPANEILEEVGVPCKKIKECEGMYMDPDPHATGALIAAAKGGAKIMNLTRVVDLILRNGGLLEGIVVVNNNTAEMAGHDVIIHVDPIALESKIVVADATGHDAVVLVELLHKRNLYNKVPNGAMWVSRSEEEVMDRTGEVYVNCFVIGLAVAAYVTPRMGPAGFSMILLSGRYGAQLIKKKLKQE

>Candidatus Nitrospira nitrosa  
MTKPRPAPLRLERDITRQIAREYYKEFDQRIESDVIIVGAGPSGLICAHDLAAMGFRTLIEQSLALGGFWHGGYLMNKATICEPANEILEEVGVPCKKIRECEGMYMDPDPHATGALIASAYKAGAKVLNTRVVDLILRRDGILEGVVVNNNTAEMAGHDVIIHVDPIALESKIVVADATGHDAIVVVELLHKRNLYQKIPNGAMWVSRSEEEVMDRTGEVYVNCFVIGLAVAAYVTPRMGPAGFSMILLSGRYGAQLIKKKLKQE

>Candidatus Syntrophoarchaeum butanivorans  
MDEVTISKAITESYMKDLIDSMLVLDVVVGAGPAGLAAAYNLAREGVKAVAFERRLSVGGMWGGMMFSRIVVQDAGREILDEIGVRCSEYEPEGYYY  
TADAEAVTTITSEVIRAGARIFNLMSVEDVVVRDDRIHGVVINWSAELSKLVDPMTVIADYVIDATGHAAEVARIVEQKLGGALTVERPM  
WAEAGEAAVENTKEIYPGLIVAGMAANAVLGS prmGPVFGM LLSGRKAAELVLSRL

>Chloroflexi bacterium RBG\_13\_51\_52  
MVKFSPVG EVVITRAIVEFAKEFN EYVESDCIIIGGGPSGLVAGRDIARAGKKVIIERNNYLGGFWSGGYLMPKVTVRPGEKILDELGV PYKT  
VAKGLVVCDA PHACAALIAAACAAAGVKIFNMTMLEDLVVKDGRVCGAVINWSPIASLPRQVAALDPVAIEAKVVIDATGH DATVVA KLEKRNL IKMK  
GEGAMWIEKSEDLIVEHT GECFPGLIVTGMAGVAYGLPRMGPTFGSMFLS GEVA AKVALEKMK

>Clostridium drakei  
MYLEDTKISKAIIDTYKDKLEDILHSDVII VGGGPGS GLV AASYLAKAGIKTTLERNS IGGGMWGGMM MNQIVIQESAKS ILDEFNIGKKYEE  
YYTADSIECVS ALTLSASQSGARILNSISVEDVIVKDKCISGLV INWAAVEKTRMPIDPIMIESKYVLDATGDASVNVNL VTRMGNVL NTPNGTLE  
GEKPMWADRGEQVIKNTREVY PGLYVSGMAANATFGG QRMGP IFGGML ISGQ KVQAELIKKIKNC

>Clostridium ragsdalei P11  
MYLEDTKISKAIIDTYK NKLED VLHSDVII VGGGPGS GLV AASYLAKAGIKTTLERNS IGGGMWGGMM MNQIVIQESAKS ILDEFNIGKKYEE  
YYTADSIECVS ALTLKAVKAGAKILN LISVEDVIEKDNCIAGL V INWAAVEKTRMPIDPIMIESKYVLDATGDASVNVNL VTRMGNVL NTPNGTLE  
GEKPMWADRGEQVIKNTREVY PGLYVSGMAANATLG GQRMGPVFGM LLSGQ KAAKQLIEKLHAAK

>Clostridium sp. JN500901  
MYLEDTKISKAIIDTYK DKLED ILYSDVII VGGGPGS GLV AASYLAEAGV KTTIERSLS IGGGMWGGMM MNQIVIQESARS ILDDFNVN YKKYEE  
YYTVD SIECVS ALTLKAVKAGAKILN LISVEDVIEKDNCIAGL V INWAAVEKTRMPIDPIMIESKYVLDATGDASVNVNL VTRMGNVL NTPNGTLE  
GEKPMWADRGEQVIKNTREVY PGLYVSGMAANATLG GQRMGPVFGM LLSGQ KAAKQLIEKLHAAK

>Coxiella sp. DG\_40  
MEQITTLGIVD SYYQKLKDNL FIDVAI VGGGPGS ALV AAYYLAKLQKVAIFERKLAPGGMWGGMM FNQIVVQSEALS ILDEFKIS YALFD NYYL  
VDSIESTASLIYHTIHAGAKVFNCYSVEDIVLKNNKVGIVVNWTGVDQGLHVDPVVAKC VIATGH SCEVAKVLA KNGIKL HETGGV VGEK  
SLAMEQAERSTIENTKEIYPGLYVCGMAANGVSGDFRMGP IFGGM LMSGK VAEIIVK DIT

>Dehalococcoidia bacterium  
MPLFHPVTEGEITRAIVNSFLRQFEEYVSSDVIIVGGGPGS GLMAGREL GKAGL KVIVIERN NYLGGFWAGGYFMN KLT LREPAQE VL DELGV PYSR  
AGEGLYVADAPHACSKLIGAAADSGV KFFN LT LEDLV VREDK RVAGAVIN WSPIA LYPL REIA ALDPVPLET KVI IDATGH DASV ARKLER RGML K  
AGE GALWIEESEEAVV EHTGEVY PGLV VTGM AVAS VYGL PRMG PTFG MLLSGK RAAEVAL AAVL TDSR

>Desulfacinum hydrothermale DSM 13146  
MALDERI IT RAIMDRYIAKLKEAIDL DVAIVGAGP SGLVAGM LLAEGKKVALFERKL SVGGMWGGML FNEIVVQEEAKTILDQVGIR AQHYTDG  
YYTADAVE SVSTL TSRSVKA GARIFNCV SVEDVM MR PERI MGLV INW SAVEMAGL HVDP LTI KARV IDATGH DSEVLH VARI KVDAELL TASG KVM  
GERSMWAEEAERL TLENT CQVY PGLYVAGMAANATFGG PRMGPIFG GM LLSGQ KVARLILEQLQS

>Desulfarculus sp.  
MLEEVIT TRAI RYLGKLDQS LEEL DAAIVGGGPGS GLVAGKKLAQAGYK TALFERKL SVGGMWGGML NEIVVQ QEARRI LEEFGV PSSE FAPGY  
YTADSVLAT STLC SVAAKAGL TIFNL VS VEDV VIRA QRTS LVIN WS AVQMAGL HVDP LTI KARV IDATGH DSEVLH VARI KVDAELL TASG KVM  
ERSLWAEQAE SDTL ANTRE AF PGV YT AGMC ANA VFGSY RMGPVFGM LLSGEK AAAEVAARLA AGE

>Desulfatibacillum aliphaticivorans  
MEERITS AIVRTYFEKLQNFLEV DLAIVGAGP SGLVAAA ALAKEGKKV AIFERKL APGGV WGGML FNEIVVQEEALH ILDDFNIS YKSAGD GLYT  
ADSVEVASGLIFGA KKAGVMINNA VASVEDV VCREGRICGV VNWT PVER LGM HVDP LVM SKA VLDGT GH PGEITDLA TRKAG KID TPTG KIM GEK  
PMW MELGEAST VENTK CLY PGLYVSGMA ANN ASGG PRMGPIFG GM LLSGEK AAAEVAARLA AGE

>Desulfobacca acetoxidans (strain ATCC 700848 / DSM 11109 / ASRB2)  
MGLDEIIISRAI TEFMEKFLDN LLELDVAIVGGGPGS GLVAGW RLAQK GRKAAIFERKL SVGGMWGGMM FNEIVVQEEAKH LL DELGITS RPY DRG  
YYTADAEIESTT LQS AMKAGV KIFNL I HVEDV M VREN RIDGLV I LWTAVN MAGL HVDP LTI RAKH VIDCTGH DVEV I KIFLR KNQ PASL KT ETGGI  
MGERSMWAEVGEAKT VEY TSEV Y PGLW VAGMTATG TLGT F R MGP IFGGM LLSGEK AAAEVAARLA AGE

>Desulfobacteraceae bacterium  
MQLDDVAISKSIL DAYFEK L LARL DV DVALVGAGP ANLVAGY YLKG SGFK A VV FESKL APGGMWGGMM FNEIVL QDDA VHIAE ELGIHCNP GGDG  
YYTMD SVEATSTIISRCV RGT VIFNL KVEDV LFRQ D RQPRV SGLV INW SPV EK LGLY VDPL SIRAS FV VDGT GH PADIC RTVARK MDV KLN TKT  
GNV VGEML WAEKGEQFTVNTA E VFP GLYVAGMAAN AFGG PRMGPIFG GM LLSGEK AAAEVAARLA AGE

>Desulfobacterium sp. 4572\_20  
MAINEV VSKAIIDR FSQKF MEY TEVDTA IVGAGP SGLIAAYFLA RAGQKVALFERQLS IGGGMWGGMM FNEIVVQ QTQ GKELLEM FG ISAREY EPG  
YYTADAVE CVTTIC SNAV KAGAKI FNC M SVEDV SIRED RV MGLV LT WSAVE AARMHV DP LTI AAKY VIDA TGH DTEVIRLIEKKADIAL QTE TGK IM  
GERSMWA EKAE QL TIENTKEI CPGV FVSGMAAN AFGG PRMGPIFG GM LLSGEK AAAEVAARLA AGE

>Desulfocarbo indianensis  
MLDEITITRAI IDRYFEKL NRN LEEL DAAIVGGGPGS GLIAGYKLAKAGYRVAMFERKL S IGGGMWGGMM FNEIVVQEEAKR ILDEVQVPTREFQPGY  
YTADSVL CT STLC SQAAGL TIFNL VS EDV M VRQ R VGLV INWTA VEMAGL HVDP LTI RAKY TIDATGH AA EVMH VARI KVDAKL FT DDG KVAG  
ERSLWAEV AETNTV NN TREAF GGF VTAGMCCNATFGSY RMGPVFGM LLSGEK AAAEVAARLA AGE

>Desulfococcus oleovorans (strain DSM 6200 / Hxd3)  
MELNEVTISRAIIDRFYEKLIANLEVDVAVGPPSGLVAAWRLARAGRKVALFERKLSIGGGMWGGAMLFNEIVVQKSALHVLDAMEIGYRLYAED  
YYTADAVEAISTLTSQAAKAGVAIFNCVTVEDVMIRPDRIVGLVLNWPVEMAGLHVDPDLMRASFVIDATGHATEVVHVVAKVPGTLRTDSGKIE  
GEKSMWSdraesitlentrevypgLyVAGMAGNATFGGPRMGAIFGGMLLSGEKVAAEILERLE

>Desulfofundulus australicus DSM 11792  
MMHLEDVVISKAIIISRYQEELLEALESDVAVVGGPSGLVAAYYLARANKKVVLFERKLSIGGGMWGGMMFNQIVIODEALPLLEEFKISYRVFEE  
GYTASSVEAVAALTLGAVRAGAKIFNLISVEDIMVRDNRVAGLVINWTPVDSLHVDPDLMRASFVIDATGHATEVVHVVAKVPGTLRTDSGKIE  
EGEKPWAARGEMATVANTREVYPGLIVAGMAANAACVCGGHRMGPVFGMLLSGQRAARIILEGDKT

>Desulfofustis glycolicus DSM 9705  
MLNEVTISTAIINRYMTKLTSALDDVAIVGGPSGLVAGYYLAKAGRKVALFERKLSIGGGIWGGMMFNEIVVQEAGAAVLAEGFLAGSPFEPGY  
YTLDHSVTTATLKHAMAAGLLIFNLIGVDDVVKDERVAGLVINWGAUTLGWHIDPLTFARYVLDATGHDAEIASVLRKMGRNLNTETGGLVG  
EKSMWAERAERETVNTREVYPGLIVAGMAANAACVCGYRMGPVFGMLLSGKRAAESILEGLA

>Desulfonatronospira thiodismutans ASO3-1  
MALEIIISRAIETYTCKLMDSELVDVAICGAGPSGMVAAYYLASAGKKTAVFERNLAPGGGMWGGMMFNEVVVQEEAREILDELDIKSVEYTPG  
YYTADSVEAVCTLGSKAAKAGARFFNLVCIEDVMIRENRTGLVINWNSAVESAGLHVDPDLMRASFVIDATGHDAEIASVLRKMGRNLNTETGGLVG  
GEKSMWAKEAEEHTIENTTEAFPGVYVCGMSANATFGSFRMGPVFGMLLSGKVAQEIIINKAK

>Desulfonauticus sp. 38\_4375  
MSLDEKIISEAIIISKYFEDFKRCLNLDVAIVGGPSGLTAAYYLAKEGFKVALFERKLSIGGGMWGGMTFNYIVVQEKGQILEEMDIICEEYKPG  
YYVDAVLATTIASKACKAGAKIFNCMSVEDVVIREEGVKRVAAGLVNVSPVEIAGLHVDPDLMRASFVIDATGHDAEIASVLRKMGRNLNTETGGLVG  
GGIEGEKSMWAEVAAEENTLKNTRAEFPGIYVCGMAANACFGSYRMGPVFGMLLSGVKAEEISTRLKEGK

>Desulfurudis audaxviator (strain MP104C)  
MKLDENIIISRAIIESYVTRLSCLEVDTVIEVGGPSGLTAAYYLARAGLKTTVYERKLSIGGGMWGGMTWNSIVVQESAKSILEDAGVALSEFKPG  
YYTASSVECAVAILTGACRAGANIMNLLTVEDVVLHNRRVSGLVLNWSAVEISGLHVDPDIATRSKFVVDATGHDSVVGVLARKAGVQLDTPSGKVG  
GEKPMWADLGEAQIMENTSEIFPGLYVVGMAANAVHGGYRMGAUTFGGMVLSGRRAEMIIDRLKV

>Desulfovibrionaceae bacterium CG1\_02\_65\_16  
MIIDERIVSEIAIASTYFGFKSCLLDVAIVGGPSGLTAAWKLAKAGRKVALFERKLSIGGGMWGGMTWNSIVVQESAKSILEDAGVALSEFKPG  
YFTADSVATAALAYQATHAGAHVFNCMSVEDVVIREVEGVKRVIGLVNVSSPVEIARLHVDPDLMRASFVIDATGHDAEIASVLRKMGRNLNTETGGLVG  
GGIEGEQSMWADVAEANTVRYTREVFPGVWAGMAANAACFGSYRMGPVFGMLLSGVKAETIDALL

>Desulfovibrionales bacterium GWA2\_65\_9  
MIIDERIVTEAIAASAYFEFKFQCLLDVAIVGGPSGLTAAWKLAEAGRKVALFERKLSIGGGMWGGMTWNYIVVQEEAKGILEEAGCAMSEYKPG  
YFIADSVATAALAYRATKAGAHVFNCMSVEDVVIREIDGEKRVGMIGLVNVSSPVEIARLHVDPDLMRASFVIDATGHDAEIASVLRKMGRNLNTETGGLVG  
GGIEGEQSMWADVAEANTVRHTREVFPGVWAGMAANATGGSYRMGPVFGMLLSGVKAEEINARL

>Desulfurella amilsii  
MALDERIISRAIERYFQKLLANIDCDCAIVGAGPAGLVCYELVKNGLKVTLFDKRLSVGGGMWGGAMMFNEIVVQEEGKILDEFDIKCSLFEPN  
YYTDSIEAITLISKTVKAGVKIFNGIEIEDVVLKKVGDYKVGVVINWTTVNMAHPVDPPIVISSFTVDAUTGHDAEIASVLRKGKVLNTDS  
GAVIGEKPMWAQIGEQDTVNHTKEIFSGLYVCGMAANAVSGAHRMGPVFGMLLSGKCAQLILEKWSRK

>Desulfurella multipotens  
MALDERIISKAIERYSQKLLSQLDCDCVIVGGPGAGLICGYELAKNGLKVTLFDKRLSVGGGMWGGAMMFNEIVVQEDGKAILDEFDIKTVLYEPN  
YYTADSIEAITLISKTVKAGVKIFNGIEIEDVVLKKVGDYKVGVVINWTTVNMAHPVDPPIVISSFTVDAUTGHDAEIASVLRKGKVLNTDS  
GGVGEKPMWADVGEQDTVNHTKEIFSGLYVCGMAANACSGAHRMGPVFGMLLSGKCAQLILEKWSRK

>Desulfurobacterium atlanticum  
MELSEVVISRAIVERFMNKLLSNLKVDAIVGGPSGLVAAYYLAKEGFKVSLFERKLSIGGGMWGGAMLFNEIVVQEMGREILDEFDVGYEKFQEG  
YYTDSVEAVTTIASKAVKAGAKVFNGBTVEDVVLKKENGDYRVCGLVINWTPVETGMHVDPDLMRASFVIDATGHDAEIASVLRKGKVLNTDS  
GCVVGEKPLWASVGEEDTVKNSREVFPGIYVCGMAANAVCGSHRMGPVFGMMSGKKIAKEIAERLKHNVEE

>Desulfurobacterium indicum  
MENLSEVKISKAIIERFTEKLLSNLEVDVAIVGGPSGLVAAYYLAKEGLKVSFERKLSIGGGMWAGAMFFNEIVVQEMGREILDEFDSVSYRKYDE  
YYTADAVEAVTTIASKAKAGAKIFNGVTAEDVVLKKVNGQYRVCGLVINWSTVDMTGLMVDPLVVTNSYVIDATGHDAEIASVLRKGKVLNTDS  
TGCVGEKPLWASVGEEDTVKNSREVFPGIYVCGMAANATCGSHRMGPVFGMMSGKKIAKEIAERLKHNVEE

>Desulfurococcales archaeon ex4484\_42  
MVKELESRVTELIVKHASRDWAELASTDVIVGAGPSGLTAKYLAEDGIKVVFERRLSFGGGIGGGMLFHKKVVEDFALDILKDFGIRYVEDGG  
LYVVDASELMAKLAVGALNAGAKITIHGVTVEDVIFRTNPLRITGVAIQWSAVPLANLHVDPDLMRASFVIDATGHDAEIASVLRKGKVLNTDS  
KSAYSELGERLVEKTKGVVPGLYVTGMAAALNNLPRMGPVFGMLLSGKVKANEVLKDLRT

>Desulfuromonas sp. SDB  
MKDINITNHIIAEFYKDIQDRVSSDVIIIGAGPSGLVASYLLAQDNFKVTVFEKRNPQGGGIWGGMMFNQLVLPDDLQDFLNQMSIKFKLHPDNL  
SVDSVFSSALLYHATEVGKIFNNIGVEDLLVVDDMVRGVVVINWNDVIKNKIPIDPLTFEAKVVDSTGHPADGVEKLARRGLVEISQEFPMNADV  
AEKFVVEATGQLYPGLYVCGMAATAAKGGPRMGPVFGMMSGKKIAKEIAERLKHNVEE

>Dethiosulfatarculus sandiegensis  
MMLDEVTITRAIIDRYMEKLNHANLDDVAIVGGGPGLVAGYYLAKKGYNVAMFERKLSIGGGMWGGMMNEIVVQEEAKRILDEFGIPCREYVEGYYTADSVSTSLTSKATLAGLSVFNLITVEDVMVRDNRVNGLVINWSPVEMAGLHVDPLTLRARYTIDATGHPAEVINVISKVDAKLSTDGKVIGERSLWAESTTIENTKEAPGVYTAGMCANAVFGAHRMGPVFGMLLSGEKVAQVLDRLQEED

>Dethiosulfovibrio peptidovorans DSM 11002  
MELDERVISKAISRSFFERLTDHLENDVVIVGGGPAGLVAAGYVLADAGVKVSLFDRLSLGGGMWGGMLFNEIVVQSEGARILDDLGVSREFEPGYYTGSVEAVSTLISSAVRAGTVFNGMVAEDVVMREDRVRIVGLVINWSTVETSGLLVDPLAVRSDFIIDATGHDSNTSTVEKKVPGRLLTETGKVEGEKSLWCERAEKLTVDNTKEVYPGLVAGMSANAVFGGPRMGPIFGGMLLSGEKAKEILLRLNGKRV

>Dissulfuribacter thermophilus  
MREIDITKAIIDKHIEELNKCLCSVVIVGAGPSGLVAGSILAQKGYTITIFEKRLAPGGIWGGGMGFKYVIQKEALDIVEFNIPYEKYSDDLYAVDAINFASGLILEAGKRGHVIFNLIAVEDLLVREGRVQGVVINNTFAKMNQFPIDPLTIEAKAVVATGHEHEVVKTLSQNDVTLNTPTGKPLGERSLFAETAEKAVVNTKEVYPGLYVCGMATAAVYGGYRMGPIFGGMILMSGKKLAGLLEEALK

>Elusimicrobia bacterium CG\_4\_10\_14\_0\_8\_um\_filter\_37\_32  
MKLDIVISKAIMETFTKDFDVYLEVDAIVGGGPAGLTAAGYVLAKKGKVVLFERKLSIGGGMWGGMMYNKCVFQEDAKKILDEFGVTHKYQEGYYVTDSLTVSILCSKAIAKGLKIFNLISVEDVMIRKEKITGLVNWNSAQLAKLHVDPMTIRAKYVIDATGHDAEVVKIVVRKIGKKLYTKGDMLGEKPMWAEVGEKDIKNTKECPGLYICGMASNAVFGGPRMGPIFGGMILSGKRISGLV

>Euryarchaeota archaeon ADurb.Bin165  
MTLDEVTISRAIITDHNLTVQYMEMDVAVIGAGPSGLVCATILAEGKLKVGLIEKKLSVGGMWGGMMFPRIVVQOGAKRLLDRFGIRSSEFSPGYYTARSIEAVAKLAAAASADVEFFNLTTEDVMVKGDGLLSGLVINWQPVEATGLHVDPMTVRCRMVATGHDAIIAHYVSKCGGLEIKGETMWADNAEAAVVAHTKEVYPGLYVCGMAANAVMGGNRMGPVFGMLLSGESAAEQILSRF

>Ferroglobus placidus (strain DSM 10642 / AEDI12DO)  
MPFSEKNITRVIREAAKEWEEISETDVVVGAGPAGLTAAYHLADFGFDVVVFERRLSFGGGIGGGMLFHKIVVEKEAKEIAEEFGIKTREVEDGLYVIDAEMLAJKLSSAGAIDSGAKVILGTVDDVYRPEPLRISGVLVQWSAVQIAGLHVDPMLIESKAVVATGHDAAVSVAAKIPPELEIYVAGEKSAYSELSKLVVEKTGKVVGDGLYVAGMAVSAYGLPRMGPIFGGMILSGRKVAEQIMFDLKK

>Fervidicoccus fontis  
MSENLEFKITKLILEHSMKDYLIEFADSDVIIVGAGPSGTAAKYLADKKLKVLVLERKLSFGGGIGGGGNLMHKIVIKSDALKIICKFEIEYKKTEFEDLYTLDASELISKLATGAINSGAKILFGYSEDIVREKPLRVSGVVWKSAIDLQALHVDPPIFTGKAILDATGHDAELIKILAKKNPSFAINVKNESSAHAELGEKQVVEFSKGKVDGLYAGMSVATLHGLYRGMGPIFSGMLISGKKVAELISKELGK

>Fervidobacterium changbaicum  
MGKDLTISKLIVENFFEKLSNALEVDAVIAAGCgpsaltslekskkgykvaifeaknepeggigwggmmfnevveleselegylkelgirffkdefiVTDVSHLASALLYHTTLAGTMIFNNVFEDLVMYDRVSGVVINWPTLREKLHVDPISIVSKFTVDTGTHPANLVKLLSKRGISSIGSTEASYNFGIVGYEFPMADAENGERFVVENTREIYPGLYIVGMAAVSVGAGPRMGPIFGGMIMSGLRAELISNELRKMGGSDDER

>Gemmatimonas sp. SG8\_17  
MRGRGGGVTDSEQQITRAIITAYHEKLWQVVGDVVVVGAGPSGLVAAATDLARRGLKTVLEKRLSPGGIWGGMAMNEVVVQDAALPLLAEFSVSRSVKGGLHVINAVERASALSILKAVQTGAVIILNTVAEDVCVHRRGRTGVVANRTNLAEALPVDPVSFEAKAVLDATGHDAAVLQMLQRGLLKGLETMQGEGPMDAAGGESFVVDKVTEVYPGLWVSGMAVVATLGGPRMGPIFGGMILMSGKRAADLISDTLSGE

>Geoglobus acetivorans  
MSYSERNITRIIVREAAKDWDISDTDVIVGAGPAGLTAAYLREFGFVVVFERRLSFGGGIGGGMLFHKIVIEEEAKEIAEGFGMKLKEVESGLYSVDSSDFLAKLSYSAVESGAKVLLGTVDDVVFDPDPLRISGVLVQWSAVQISGLHVDPMLIESRAVVATGHDAAEVISIAARKIPPELEIFIHKKSAYSEMSEKLVVEKTGKADGLYAGMAVAAVHGLPRMGPIFGGMILMSGKKVAEQIMFDLKK

>groundwater metagenome  
MIDETKITELIVRSAVDDFLGNLKVDVVVVGPGAGLTTARYLAKAKRVLFERKLSIGGGMWGGMMFPRVVLQKGGKEILECNVRYKKFDDLWVADSIECVTKMTAKAIDEGVKIFNLISIEDVIIRSVQSNKNNKEGKTKICGVVNLWTAQMANLHVDPLSVKSDFVVVATGHEASICHLVVKVGNLNTKTDVGLERSMWAEGKESDIMNNTKEVYPGLVSGMAANAVYGSERMGAIFGGMLMSGKKVSELILEKEK

>Hadesarchaea archaeon DG-33-1  
MGGIEDTEITAIIKRFMRDFEDVTNLDAIAAGAGPSGITAASFLASGGAKVAVFERNLHVGGMWGGGILFSRVVIQEAKVMLEEVGVKLKPTAAGYYTADSVEAVTKSTAAVDAGARVMVGLTVEDVMIREKDRVAGIAVNWKAVELAGLHVDPVGISAKIVIDATGHDAMIRIVQRKVPNAKFPTSTGGVGEKPVWAEGEIEVNNTREIYPGLIVTGMAANTVFGSPRMGPIFGGMILSGRRAEVALKV

>hydrocarbon metagenome  
MWYLVLELDERSRAIIAVQMEKMLRYTDMDAIVGGGPAGLTAASFGLAEGFSVALIEKKLSVGGMWGGMMFPRIVVQEEGRQLLDHFAIRYTRYEETYVASSVEAVAKLTAACDAGVEFFTIVTVEDVMVRSDKRLSGLVITWSPVEMAGLHVDPMTLGCRTIDATGHDAIVARLVARSGAVTVKGEGFMWADRAESRITSHTREVFPGLIVAGMAANAVAGENRMGPVFGMILSGRHAALVSRELASKP

>Hyperthermus butylicus (strain DSM 5456 / JCM 9403 / PLM1-5)  
MVNAVQAPEHSWLPHNVTSLREGALAALIIRKTAEKLTSITSVDVAIAGAGPAGLTAAWLLAEKGLRVVVVEHSLGVGGMRRGSGMLMPVGLVEDGPPAELLRRAGARLDRVADGLYAVDPTEAVVKAIAKIDAGAVILPGIHVEDLILWRSGSGYRVAGLVINLSPVVEAGWHVDPPIYIEARATIDATGHDAELVKLLSKALGDSSIRVRGTRGMDVWEGEKLVVETGEVYPGLYAGMAVSETYQLPRMGPVFGMILASGARVAELVASRLSEQ

>Ignicoccus islandicus DSM 13165  
MIDEVKVTSIIIEESSKELSMAKGVDVVIVGAGPAGLTASHYLAKAGLKVLILERRVSLGGISGGGSIFHKVVVEDVELEGYNPKEIAEELGVPLKKVDDNLYTTDAALVAKLSNASVSAGAKIVLGMHVEDLIYRIEEGVTKVKGVALWSPIYLSGLHVDPPIFFKAKAVVDTGHDAEILKIAKKLPNVNEFVGREYGAVIDAEKLVVKYTGKVLLEGLYAAGMSVASFYRLPRMGPVFGGMLASGKKVAEKIIGDLEVS

>Ignisphaera aggregans (strain DSM 17230 / JCM 13409 / AQ1.S1)  
MKELELRISRAILRNSVRELIEYSDV ррр VV ррр VV GAGPSGLTAARYLAMNGFRVVVLERRLSFGGGIGGGGMLFH KIVSSEALPILNDFDIKYRDEEDLYMIDSSELMAKLAVGAINAGAKIFHGHIHVEDVIYRENPLRITGVVIQWSAVVMSGHLHVDPFLITSRAVVDTGHDAEVLQIVSRKIPEVGISLPGESSAYSELSEKIVVEKTGMVIPGLYVAGMAVAALYKLPRMGPIFSSMLLSGKRVAEIIANDLKK

>Korarchaeum cryptofilum (strain OPF8)  
MESLESRISKAIWESTYKDLDIIDS D VV VV GAGPSGLTAASYLAKSGFKTTVIERRLSFGGGIGGGGMLH KVVVDGRALKVLEDFKVRYSTYLEKYDLYVLD S AELMAKLASGAIDSGAKLH GLTVEDLIVREDPFRVEGVVVQWSSVLLAGLHVDPFLIHSRVVVDTGHDAEVIRILERKNPSLGIKVPGERSAYSELSELSVVERTGKVEGLYVTGMAVAALNQLH RMPGPI FSGM LLSGRKVAEEIIRDLS

>Labilibaculum filiforme  
MEQIVSAGIVDSYF KKLKENLS DVAIVGGGPSGLVASYYLAKKGFKVALYESKLA PGGMWGGAMMFNEII VQKDALHILNELGVSYQHYQEDYYTLD SVHATSALIYHATQAGVKIFNC S FIEDVV FQNDKVC GVVL NWSPVRREGLHVDPFLVVMAKAVV DGTGHDCIARTLERKNDV KLN TKGK VM GEC SLSIDEAERTTVENTKEIYPGLYVSGMASNGVSGGFRMGP IFGGMLMSGKKAELTI DKPAPKRCLPDE

>Latescibacteria bacterium DG\_63  
MKLDDVEISKAIIESFYAKLDSLMDV ррр VV VV GAGPSGLSAAYYLAKKGLKTAIFEKSLAPGGVWGGMLFNEIVVQENVLGILEQIGISYQAVPAKGYTVDSVEMASGLIFNAVKAGAKIFNAMSVEDIVFKEGRVNGLVINWSPVRKLAMPDPLTVIAKAVVDTGHPC EII RIACEKAQVKIATETGGV LGERPMWVQHGEQQTV DSTA EYY PGLFACGMSAT NVTGGYRMGP IFGGMILMSGKKAADLIAKSLASLEKR

>Metallosphaera yellowstonensis MK1  
MEIRQDEVKITKYILKATFEDWMEDIAENDVV VV GAGPSGLSAAYYLAKKGLKTTVFERRLSFGGGIGGGMLFH KIVIESPADQVLREM NIRLQ RV EEGVYIVDSSEFMAKLASSAIDAGAKIVHGVTVDDVIFRENPLRVTGVAWEWTQMASLHVDPFLI HAKAVVDTGHDAEVISVAARKIPELGIA PGEKSAYSEVAEKLTV DNTGEVAPGLYAA GMATTEVKGKLMRGP IFGGAMVLSGKVAEDIASTLLMKARNT

>Methanobacteriales archaeon HGW-Methanobacteriales-1  
MELDDITISRAIVEFMND MFDYMDIDVAI GGGGPGAGL TAGYYLAKAGLKVALYERKLSIGGGMWGGMMFNKIVVQEEGKRILDEFGIQS SKK YQEN YYVSDSVEATSTLC SKATQAGLKIFNLMSIEDVMIRGDDISGLV INWSSVEMGGLHVDPFLSIRSKAVIDATGHPCEVV KVVQNKIGPKLNPTGEII GEKSMWAEVGEPAIMENTREVYPNLYVAGMAANAVYCAPR MGPVFGGMLLSGEKIANMLIEKLK

>Methanobacterium subterraneum  
MKLDIIVSKGIVAGYMEELLDYMDMVAI GGGGPGAGL TAGYYLAKAGLKVALFEKKLSMGGMWGGMMFNKIVVQEEGKRILDEMGIRNQ EYE EG YYLADSVESASTCSKACQAGLKVFNLMEIEDVMIKGEVGLVINWSPVEMAGLHVDPITVGARAVIDATGHPCEVV KVL ERKMEAPLKTETGKIM GEKSMWADVAEQNIMGNVGEIYPMGVYTGMAANAVHGSQRMGP IFGGM LLSGEKVAEMLIEKLK

>Methanobrevibacter woesei  
MKLDDITVSKAIIQEYMNDFLDYTDMDVAI GGGGPGAGL TAGYYLAKAGYKVALFERKLSIGGGMWGGMMFNKIVVQEEGKRILDEFGIQS KKFED NYTVDSIECTSTLC SKATQAGLKIFNLMSIEDLMVRENGINGIVLNWSSVEMSGLHIDPLTVRAKAVIDATGHPTEITKIVEQKMGANLKTETGKIM GEKSMWADRAEGKILDNVTEVY PGLWVTGMAANAVHGSQRMGP IFGGM LLSGEYVAQKII EKLENE

>Methanocalculus sp. 52\_23  
MQLDEVTISRAILETHAEISSRYL D DIAIVGGGPSGLVCAALAAAEDGRKVAVIEKKLSVGGMWGGMTFPRIIVVQEEGKRLLDQFGIRSRVYKPG YH VASSVESVAKL TAAACDAGAEFFNLT SVEDVV I KEDGRVSGL VIT TSPVEMTGLHVDP LT LAKVTV DATGHDAVVAHC VL RGKDITIHGESFM WAERAETNIINH TREI F PGLIACGMAANAVAGEARMGPVFGGMLLSGEHA AVLAREISERV

>Methanocella conradii (strain DSM 24694 / JCM 17849 / CGMC 1.5162 / HZ254)  
MELDETLISRAIIIDFLRTLSDYVSDVGIVGGGPSGLVCATYLARAGVKVAVFERKLSVGGMWGGMMFPRIVVQEEATRILDDFGIRYREYRPG YYIAGSIEAVGRLT SAAAGAGAEI FNLM SVEDVMIRENKEVGLVINWSAVDIAGLHVDP LTRV VV DATGHPAEVCRIVERKVSGGAFKVPGEQ SWADRGERALISTTKEVY PGLV VAGMAANAVAGGPRMGP IFGGM LLSGEIAARIVKEKLGVS

>Methanococcoides burtonii (strain DSM 6242 / NBRC 107633 / OCM 468 / ACE-M)  
MKLDEVTISRAIIIEFSKVFLDYTDV ррр VV VV GAGPSGLVA AKYLA EAGLKTVIYEKKLAVGGMWAGGMMFPRIVVQEEATRILDDFGIRYREYENG YYVANSIESV GKLI SGAT SAGAEI FNLM SVEDVMIRENKEVGLVINWSAVDIAGLHVDP LTRV VV DATGHPAEVCRIVERKVSGGAFKVPGEQ GEKPMWADVGKML DTTKEVY PNL YVAGMAANAVAGA PRMGPVFGGMLLSGEIAARIVKEKLGVS

>Methanococcoides methylutens MM1  
MKLDEVTISRAIIDEF SKVFLDYTEVDV ррр VV VV GAGPSGLVAA KYLA EAGLKTVIYEKKLAVGGMWAGGMMFPRIVVQEEARHILDDFGIDYHEYENG YYIANSVESV GKLI SAAT SAGTEI FNLM NVEDVMIRENKEVGLVINWTAVEI GRLHVDP LTRV VV DATGHPAEVCRIVERKVSGGAFKVPGEQ GEKPMWADVG ERM LVETT REVY PNL YVAGMAANAVAGA PRMGPVFGGMLISGKQVAELI IERL

>Methanococcoides vulcani  
MKLDEVTISRAIIDEFSKVFLDYTEVDVALVGGGPANLVAALKYLAEGALKTIVIYEKKLSIGGGMWAGGMMFPRIVVQEEARHILDDFDITYHEYEKGYIANSVESVGKLISGATTAGTEIFNLVNVEDVMIRENDEVGLVINWTAVEIGRLHVDPLAIRAKVVVDTGHEAAVCNTVQRKVPGAKLGDLGVVGEKPMWADVGERMLLETTKEVYPNLYVDGMAANAVAGAPRMGPVFGMLLSGKQVAELIIERLK

>Methanococcus maripaludis  
MDGKLRADEVAVTKSILKSTFDMWMDLIDVDVVIVGAGPSGLTAALKYLAQNGVKTVVLERHLSFGGGTWGGGMGFPNIVVEKPADEILREAGIKLDEVIGEPELTADSVEVPAKLGVAIDAGAKILTGIVVEDLILKEDKVSGVIQSYSIEKAGLHVDPITISAKYVIDSTGHDSSVIHTLARKNKDLGIEVPGKSMWADKGENSENLTRNTREVFPGLYVCGMAANAYHAGYRMGAIFFGMYLSGKCAELILEKLENK

>Methanocorpusculum labreanum (strain ATCC 43576 / DSM 4855 / z)  
MDLEVTKAITESWFARLQENLCFDAIVGTGPGSLIAAVKLADAGYKVSMFESKLAPGGMWGGAMLFSSIAVQNEAVYLLDELEIPYKRYNENLVVCDSDLATSALIYQASKRGVVIHNGMSVEDVVFMDNRVSGVNVNWGPVREGLHVDPPLSFRAKIVDATGHPCMISETAARKNNITLNTPTGKVCGEC SLNAVEGEAMTVENTKEIYPGLYVCGMAANGVFGSPRMGPIFGGMLLSGEKVAKLIIIEKL

>Methanoculleus thermophilus  
MTINEVTISRAILESHRALIEHLEMDDAVVVGGSGLACAALLGEKGLSCALIEKKLSIGGGMWGGMMFPRIVVQEEARRLLDRFGIAYKEFEPGYYVAKSVEAVAKLTAACDAGVEFFNLTTVEDVMIRGDGRVGGLVINWTPVDMAGLHVDPPLTACTCTVDASGHDAVVARMIERKGGLQVKGESFMWAERAESRILDHTKEVFPGLVAGMAANAVAGECRMGPIFGGMLLSGERAELVAESLER

>Methanohalobium evestigatum (strain ATCC BAA-1072 / DSM 3721 / NBRC 107634 / OCM 161 / Z-7303)  
MELDDITITKAIVDDFSKTFIDYTEDVALVGGGPANMIAATRLAQEGYKVALFEKKLALGGMWGGMMFPRIVVQDEARKILEEFDINHYEYDNEKGYIANSIESVSRLINKTVTSVQVFNVNFEDVMIREDDRVTGIVINWTAVSIANLHVDPPLTIRAKVVIDGTGHEAVVCNTVQRKIPNAKFEGVVERPMWADAGEKSLLKETTREVYPGLIVTGMAANAVAGAPRMGPVFGMLLSGEMAAKIAMSCLD

>Methanohalophilus euhalobius  
MELDERIITRAIVEFTNVFLDYTDVDFVALVGGGPANLVAARYLAEGALKTFLFEKKLSVGGGMWGGMMFPRIVVQEEARRILDDFDVPHYEYEEGYYVANSVGTVGKLISAAVSAGVEIFNLVSFEDVMIRDNEVCGLVINWTAVEIARLHVDPPLTIRAKVVIDGTGHEATVCNTVQRKIPGAFFGKEVVGEEKPMWADTGERLVMKNTREVYPGLIVTGMAANAVAGSPRMGPVFGMLLSGEKAQALASRLKD

>Methanolacinia petrolearia (strain DSM 11571 / OCM 486 / SEBR 4847)  
MKLDEVTISRAILSEQHKIMTEYLDIDCAVVGGSGLTCAAILQAQNGVKVALIEKKLSIGGGMWGGMMFPRIVVQEEARRLLDHFGIKYTEYEKGYYVASSVEAVSKSLAAACDAGAEVFNLTTVEDVVVKEDGGVSGLVINWTPVEMAGLHIDPLTMRTKTVDATGHDSMIAHMVRKKGAALEIKGEGFMWAERAETNILSHTKEVFPGLIVAGMAANAVGGETRMGPIFGGMLLSGEKAANMIIERLK

>Methanolinea sp. SDB  
MELSETTIITRAIVSSQMKİLEYSELDAVAVVGAGPSGLTAAAILGDAGYKVGVIKKLSVGGMWGGMMFPRIVVQEPAARRLLDRFEISYQPFEEGYYVASSIEAVARLTSAACRGAEFFNLTSVEDVMVKDDGRVSGLVINWTPVEMAGLHVDPPLTICRYTIDATGHDAVVATLVERKGRNLEVKGEGFMWADRAESEIISHTREVYPGLIVTGMAANAVAGEHRMGPVFGMLLSGEFAASLVREKLNR

>Methanolobus profundi  
MELDETIITRAIVEEYSKVFPLYIEVDVALVGGGPANLVAALKYLAEGALKTFLFEKKLSIGGGMWGGMMFPRIVVQEDAKHILDDFNINYHEYEKGYYVASSIESVGKLICGATDAGAEIFNLIDVEDVMIRENDTVCGLVINWGPVSMNRLHVDPPLAIRAKVVIDGTGHDAGICSTVQRKIPGTDIKLDVVGEEKPMWADVGEKILMDTTKEVYPGLIVTGMAANAVAGAPRMGPVFGMMLSGKKAELAIEKLRK

>Methanomassiliicoccales archaeon PtaB.Bin215  
MEIDEVLVLTRKIVERYTEEFLENVDVVIAAGAGPSSLTAARYLAKAGLRVVIKERKLTPGGMWGGMTFPIIVVQEGSKDLLGEIGVRLRDAGDGYFTADSVEASAKLISAAVTAGARLYNTISVEDVMIRQDSICGVVINSSAVEAGLHVDPPLAVRSKYVIDGTGHPAEEVHHVVKVGRNLNTPTGQIEG EKSMWAEQGEKDVVELTGEVLPGLYVTGMANAVAGAPRMGPIFGGMLLSGRKVAEMIAREKKKGKK

>Methanonatronarchaeum thermophilum  
MNVDKFKVSKAIIDEFSKDFLDSLSDVDAIGGAGPAGMVAALKYLAENDIKTAVFERKLTSVGGMWGGMMFPRIVIKEKSLPILDDLNNINYREYQDGYYIANSIESVGQTAAEAVKAGAEIYNLMTVEDLHYKENKVNGVVINWSSVLDAGLHVDPPLTIESKITIDATGHDCELVKVAQERINKLNKTGKIMGEKSMWAEQGEKDVVELTGEVLPGLYVTGMANAVAGAPRMGPIFEGMLLSGKVAEQCICKLK

>Methanoplanus limicola DSM 2279  
MKTKVYESENKMLDEVAISRAIVSEQSVMFLYYDDLCIAIVGAGPSGLTCAMLGEEGLKVGVIKKLSVGGMWGGMTFPRIVVQEEARRLLDHFGIKYREYESGYFVSSSVEAVAKITSAACDAGAEFFNLTYVEDVVIKGDNRISGLVINQTPQMTGLHIDPLLATKVTIDATGHDSVVAHLVRDKGGSVEIKGEGFMWADRAESNILSHTKEIFPGLIVTGMAANAVGGETRMGPVFGMLLSGEKAALKSALKK

>Methanopyrus kandleri (strain AV19 / DSM 6324 / JCM 9639 / NBRC 100938)  
MEREITPIVLREGYEFINDCESDViVVGAGPAGLTCAYELAKSDVDTIWERKLYVGGGMWGGMLFPAGVIMEETAEVLEEVGVELRPAEAGLLAFNVEAAIKLANALEAGARILVGLIEVEDVIERRGRVCGVNVNWTAVKAANMHVDPLALEAEYTVDATGHEAAVCKLAGIEVKGEPMWAERGEELVKHTQEVKPGFLVAGMAASAVKGAYRMPGIFGGMLESGKAAEELERL

>Methanoregula formicina (strain DSM 22288 / NBRC 105244 / SMSP)  
MELDELTISRAILASQTNVLINHLEDDAVVGGGPAGLTCALIAGQGKKVGVIEKKLSVGGMWGGMMFPRIVVQEEARRLLDGFIRYTFESGYYVARSVEAVSKLTAACDAGVEFFNLMSVEDVMIKADKRISGLVINWTAVEMGKLHVDPLOVMSRTVDATGHDAVVARLVEKKGGDIRVKGEKGFMWADRAETNILNHTKEIFPGLVVAGMAANAVAGESRMGPVFGMFLSGERAQIVLREMKA

>Methanoregulaceae archaeon PtaB.Bin009  
MELDEITISRAILSSQVEKLLFEMDVAVGPGPSGLTAAALIGEKGFRVGLIEKKLSVGGMWGGMMFPRIVQEEAKRLLDQFDIAHTSYEGYYVASSVEAVSKLTASACDAGVEFFNLFSVEDVMIRGDSRLSGLVNNWTPVEMAGLHVDPMTGCRAVDATGHDAVLARLVERKGGDVKRGEGLWWADRAESEIVSHTREVFPGLVVCGMAANAVAGEHRMGPVFGGMLLSGERAALATSSLRQENSAA

>Methanosaeta harundinacea  
MALDEVITITKAIVESYMESFLKYTDVDVALVGAGPANLVAAKKLAEADAKTVVFERNLSVGGGIWGGMWGGMMFPRIVQKEGCRLDEFGVWYREYEGLYYIASSIETVAKLTAGVIDAGAEIINLVTVEDVMIREDERIALVINWEAVERTRLHVDPMSVRARVVIDGTGHDANICKVVQRKIPGAKVGSGLGVVGEKPMWADVGEKTVVETQEVYPGLATGMAAAAAGVAGGPRMGPIFGGMLLSGEKAALALEKLGL

>Methanosalsum zhilinae (strain DSM 4017 / NBRC 107636 / OCM 62 / WeN5)  
MELDEVVITRAIVEDEFNLNVFLDYTDVDVALAGGGPANLVAAKYLAEAGYKTVLFEEKLSSIGGGMWGGMMFPRIVQEEARRILDDFNITYKEYEDGYYVANSIESVSKLAAGATSAGAEIFNLVSEDVMIREDRVSGLVINWTAVGIGKLHVDPLTIRSKVVIDGTGHDASVCNIVQQKVPGAQLGELGVVGEKPMWADVGEKLLMETTREIYPLIVSGMAANAAAGAPRMGPVFGGMLLSGEKAELAISKLD

>Methanosarcina acetivorans (strain ATCC 35395 / DSM 2834 / JCM 12185 / C2A)  
MELDEVIITRAFDEYSKTFDIDVALVGGGPANLVAAKYLAEAGVKVALYEQKQLSLGGGMWGGMMFPRIVQEEATRILDDFGIRYKEYESGYYVANSVESVKGKLIAGATSAGAEVFNLVSFEDIMIREDRVTGIVINWGPVTTQRLHVDPMLIRTKLVIDGTGHEAVVCNTILRKIPNAKIGELGLLGEEKPMWSEVGERLAVNATQEIYPLIVAGMAANAA TRAPRMGPVFGGMLLSGEKAALKALDLRKLT

>Methanospirillum stamsii  
MTLDEITISRAIIISDMHMTLEMEMDVAIVGGGPSGLVCALSIAEKGYKVGLIEKKLSIGGGMWGGMMFPRIVQSEAKRLLERFNITHSEFSPGYYTARSIEAVSKLTTAAVDAGVEFFNLTTVEDVMVKGDGRSLGLVINWQPVEATGLHVDPLTIRCRMIVDATGHDAVIAHYVSKMGKPDIKGEGTMWADNAESAVVTHTKEVFPGLVFCGMAANAVSGGHRMGPVFGGMFLSGESAAVQILQQL

>Methanothermobacter defluvii  
MEIHAGVKMKLDDIKISRAIVEGYMEDLLDYMEMDVAIGGGPSGLTAGYYLARAGLKVALFERKLSIGGGMWGGMMFNKIVVQDEGREILDEFGIRSEPYDEGYHVADSVEATSTLCSCAQAGLKIFNLMISIEDVMIRDEGITGLVLNWSSVEMAGLHVDPVTRAGAVIDATGHDCEIVKVVERKIGPELNTPDGRIQGERSMWADVGEEAALIENTREVYPNLVYAGMASNAVYGA PRMGPIFGGMILVSGRRVAEMIEKLK

>Methanothermococcus okinawensis (strain DSM 14208 / JCM 11175 / IH1)  
MDKFKIEEKDVTTISILKATFNMWMDIVDWDVVIVGAGPSGLTAARYI LAKEGVKVVVVERHLSFGGGTWGGGMGHPYITVQKPADEILREVGVKLEEIDGGLYVADSVEVPAKLGVGAIDAGVKILTGVIVEDLILKENKVSGVVINSYAIKDAGLHIDPLTINA KYVIDATGHDASVTNLARKNDLGLGEVPEGKSLWAEKAENSILRH TREIFPGLVFCGMAANATHGGYRMGAIFGGMYLSGKKVAELILEKLKNND

>Methanothermus fervidus (strain ATCC 43054 / DSM 2088 / JCM 10308 / V24 S)  
MVLNEVTISKAIISKYMEELIDNTNLDAVIAAGGGPSGITAGYYLAKEGFVKVALFEKRVSVIGGA WGGMMFNKIVVQEEGKKILDEFDVNTERYENNYYVADAIEMITTLASKACKSGLKIFNLINE DIVINKKKISIGVVNWTAEMA KIHVDPLVIKSFKVIDATGHDCEVVKA VEKKLGPVLTETGRIVGEKPMWAEKGEKAVIKNTGEVYPNLVYAGMAANSVYGSYRMPGIFGGMLLSGKKVAELIRERLL

>Methanotherrix soehngenii (strain ATCC 5969 / DSM 3671 / JCM 10134 / NBRC 103675 / OCM 69 / GP-6)  
MSLDEVMVTKAIVEGYLESFLENTEVEAALVGAGPANLVAAKRLAENI KTVLFEKRLSLVGGWLWGGMMFPRIVVQQAIRILEEYGYRHEHCKGYYVANSIETVAKLTARAIDAGAQIVNLTVEDVMIREQDRVVGVLVINWTAAEMAQIHVDPLCIRARYVIDGTGHEASVCRVVAR KIPGAI GIGIDGVVGEKPMWAEVGERTV VEMTQEVYPGLVVGMAAAAVCGGPRMGPIFGGMLOQSGEKAAGIVIENLNK

>Methanotherrix thermoacetophila (strain DSM 6194 / JCM 14653 / NBRC 101360 / PT)  
MALDEVKITRAIVESYLESFLKCTDVDVALVGAGPANLVAAKRLAEADVRVVLFEKRLSLVGGWLWGGMMFPRIVVQKEACRILDEIWIYREFEEGYYVADSIEVVAKL TAGAIDAGAE LINLVSEDVMIREGDRIVGLVINWTAADMAGI HVDP LAIRARVVIDGTGHEASVCRVVAR KIPGAI VGESGVI GEKPMWAALGEKIVV DATREVYPGLIVAGMAATTVAAGPRMGPIFGGMLLSGEKAASIALEKLQAQSVD

>Methanotorris formicicus Mc-S-70  
MDLRLKADEYTTKAILKSAFNWMDDIDVDVAIVGGGPSGLTAARYIAKKG YKVVLERHLA FGGGTWGGGMGF PYIVVEEPADEILREVGKILEKVDGEEGLYTADSV E VPAKLA VGSIDAGAKILT GIVVVEDLILREN R VAGV VINSYAI EKAGLHIDPITITAKYVVDATGHDASVATL SRKNPELGLEVPEKSMWA EKGEN ALLRNTRREVYPGLVFCGMAANATYGGNRMGAI FGGMYLSGKKCAEMVVEKLNNE

>Nitrospira bacterium SM23\_35  
MELDEVVITKAIVDQFCKKLTNHLET DVAIVGGGPSGLVAGYFLAKAGRKTVLFERKLSVGGMWGGMFLNEIVVQKA VRILKEFGITYCEFQKNYYTADSVESISTLISRAVQAGVTIFNCITAEDVLMRTSRTGLV LWNWSAVE MARLHVDP LA VR SRFV VDATGHETAVV RL VQN KVP GTL KTL SGK VGEKSMWSDKAES TLK NTREVFPGLVYAGMAANATFGGPRMGPIFGGMLLSGEKA KL LQAL SQKSK

>Nitrospira japonica  
MGKPKPAPL RERDITRQIAREYYKEFDQ LIESDVIIVGAGPSGLICA HDIADM GFKTVIVEQNL ALGGGF WHGGYLMNKATICAPAHK ILDEIDVPC KRIKDCEGMYIVD PHATG ALIA AAYRAGAKVNLTRV D LILR RDGSLDG VV VNN TAE MAGHDLI HVDP IALES KIVV DATGHD A VV VNL LH KR SLYTEVPGNGAMW VSRSEEDVMDHTGEVYPNCFVIGLAVSAVHGTPRMGPAGFSMLLSGRYGAELKKKLQK

>Nitrospiraceae bacterium  
MNPINVAPL RERDITRHIAREYYKEFD S LIESDIIIIVGGGPSGLL CARD LATSGF RTLLIEQSLALGGGFWSGGFLMNKATICEPADQI LEELGIPFKPIKDCPGMTMVDPPHVT SRLISAAYEAGVKIMNLT KVV D LILR QDHRIEGVVVN STVEMAGHDTI HVDP IALES QIVV DATGHD A VV VNL LH KR NLQYK VPGNGAMW V ARSEALV VENTREIYPNCFVAGLAVA AVDGS PRMGPAGFSMLLSGRYAA ELVR QKLGE

>Nitrospirae bacterium  
MPKPPTPAPLRLERDITRHIAREYYKEFDQLIESDVIIVGAGPSGLICAHDLAAMGFKTVVVEQSLSLGGFWSGGYLMNKATICEPANEILEEIGVP  
CKKITECAGMMYMDPPHATGALIAAYRGAKIMNLTKVDLIIRRDIILEGVVVNSTTAEMAGHDIAHVDPPIALESKIVVDAKGDAIVVELLHKR  
NLHKAVPGNGAMWVAQSEQEIMDRTEVYVNCFVIGLAVAAYGTPRMGPAGSMSLLSGRYGADLIKLLKG

>Omnitrophica bacterium RBG\_13\_46\_9  
MDEALISRAITESFTKDFIDAFNVDAIAAGAGPSGLICACYYLAKQNWKAVFERHLRVGGGMPGGMMFNRIIVQEEAMPILEFGVSAKRYKKDLY  
IVDALEAISTFCSTIKRGAKIFNLINVEDVVIRKDRIAGVVLNWSAVSWAKLHVDPMAVRSKAVVDAKGHDSEIARIVERKTGPVLRTEGGVIGE  
KSMWAEIGEKMILENTKEIYPGLIVCGMAANAVFGSPRMGAIFGGMLLSGKAAEVARKVIKSK

>Omnitrophica WOR\_2 bacterium RIFCSPHIGH02\_02\_FULL\_68\_15  
MFARAQEAQITRAVRAFAKEFDGLVRSVLIVGAGPSGLVAAMDIARRGRRLVVEQTNYLGGGLWLGGYLBNKLTVRAPAHRLKELKVP瞿Q  
PGLYVADAPHVCARLIAAACDAGVKFAQMTEMVDVVVREGGRVEGLVINWSPVSALPKLAHVDPVALEAKVVDAKGHDAAVVRLLAKRGLAAPVP  
GDGAMWVERGEQAVMDKTGEVHPGLFAAGLAVSAVHGTPRMGPAGSMSLLSGRCAQMIERAYFA

>Peptococcaceae bacterium SCADC1\_2\_3  
MPLEDITISKAIITTRYNQELLALESVDVAIAAGGGPSGLVAASYLAQQGAQVVLVERNLSLGGMWGGMMFNQIVVQEEAIPILDTGFVRYRTFEPG  
YYTAHATEVAAIILGAVRKGVKILNLISAEDVMVRNERVCGLVLNWTAVGALARLHVDPIAASCSCVIDCTGHDAQIANIVRKMGAVLKTSRGKIE  
GEKPMWAERGEAAIKNTGEIYPGLVAGMAANAVGNHRMGPVFGMLLSGKRAELIKGER

>Phorcysia thermohydrogeniphila  
MQNLNEVIISQAAIESFMEKLKNSLEVDVAIVGGGPSGLVAGYYLAREGFKVSIYERHLAIGGMWAGGMLFNEIVVQEMGREVLDEFGVRYREFQP  
GYYVADSVEAVTTIASKAVKAGAVIFNGVTAEDVVLKKVNDEYRCGLVINWTSVERSRLPVDPVITAKEYVIDATGHDAVSVSTLQKKAGIKLATE  
TGCVIGEKPLWASVGEEDTVKNTREVFPGLFVSGMAANATCGSHRMGPVFGMLVSGKAAQEIAEKLLKGKNEE

>Planctomycetes bacterium DG\_20  
MDDDFDETDSQAILRAYYAKVADALQGDVLVVGAGPSGLVAAWRLAQAGHRVVVLEKRLSPGGGIWGGSLGMNEVAVQKHIALDEAGVRHQPSGR  
LFTADAMELASCLKALHAGAVILNLMTAQDVCRSGRTGVVANRSLLGESLPIDPIVSARAIDATGHEAVLANCIQRGLLKNSLGLPGEGL  
PLDAPAGERFVVDHVAELYPGLWTGMSVCASLGPGPRMGPIFGGMLLSGEKVAALVGQALKTRPQVRHE

>Porphyromonas sp. CAG:1061  
MEKLVSQGIITTYFEKMEKSLDLDVAIVGGGPSGIVAAAYYLAKAGLKVAQFDRKLSPGGMWGGAMMFNEIVIQQEEALEIIKEMGINYEQYQDKLYT  
MDSVESTALLYNAVHAGARIFNCSVEDVVYKENRVSUVVNWTPVLRGMDPLNIMAKYVIDGTGHDSEICRVVAKKNGATLNTSTGGVGEQ  
SLDVITGEKMVEGTKEIYPGLYVCGMASSAVGTTPRMGPIFGGMILMSGKKVANLIIDQLK

>Prevotella amnii  
MIEKEISKGIITTYFEKMEKSLDLDVAIVGGGPSGIVAAAYYLAKAGLKVALFDRKLSPGGMWGGAMMFNFQIVIQKEALDIKEFEINYEQYSDNLF  
TTDSIECTAAILYKAHVAGATIFNCYSVEDVVFKNNIVSGVVVNWTPVLRGMDPLNIMAKFVIDGTGHDSEICKVVARNNITLNTSTGKVVG  
RSLDVIEGEQQVVEGSKEIYPGLYVCGMASSAVGGTPRMGPIFGGMILMSGKKVADMILKRIQS

>Prevotella nigrescens  
MIEKKISKGIITTYFEKMEKCLELDVAIVGGGPSGIVAAAYYMAKAGLKVALFDRKLSPGGMWGGAMMFNFQIVIQEEALEIIKDFDINYQAFEDGLY  
TADSVESTSALLYKATHAGATIFNCYSVEDVVFKNNIVSGVVVNWTPVLRGMDPLNIMAKFVIDGTGHDSEMCQVVARNGIKLNTATGDIV  
RSLDVAEGERQVVEGTKEIYPGLYVCGMASSAVGGTPRMGPIFGGMILSGKKVADMIILERLK

>Prevotella stercorea DSM 18206  
MIETQVSKGIITTYFDKLQNNDLDVAIVGGGPSGIVAAAYYMAKAGLKVAQFDRKLSPGGMWGGAMMFNFQIVIQEEAMHIVKDFDINYQAFEDGLY  
TIDSVESTSSLLYHAVHAGATIFNCYSVEDVVFKNNIVSGVVVNWTPVLRGMDPLNIMAKCVIDGTGHDSEMCVVARNGIQQLDTATGGVIGE  
RSLDVVEGERMVEGTREVYVPGLYVCGMASSAVAGTPRMGPIFGGMILMSGKKVADMIIEKLKK

>Prosthecochloris sp. ZM  
MEEKISKFIISQFFAKLEDSTLVDVAIVGAGPSGLIAAKELAKAGKVAIFESKLAPGGVGWGGMLFNEIVLQENIIPILDEYAIRYKTGEFYVT  
ADAVEVSSALIYGAHVAGVRIFNAVRVEDLAMRDERVCGVVINWNPVSRLLEMHVDPVITSRAVLDTGHFSELINLASNKAGITLDTPTGKVMGEK  
PMWMENGESSTVINTKRLYPLGLYASGMAANNAMGGFRMGPIFGGMILSGKKVAGLILEDIQG

>Pseudothermotoga lettingae (strain ATCC BAA-301 / DSM 14385 / NBRC 107922 / TMO)  
MKDTMISTLIVNRYFKKLRSFLELDVAIVGAGPSGLTAAYELAKKGKVAIFEEKNTPGGIWGGMMFNEIVLEKELEDFLNELGITYVIQENHVL  
VDSVHFASALLYRTTMVGATVFNNISVEDVAMQDGKVCGVVINWGPTMRLGLHVDPITVKASFVIDGTGHF PANVASLLAKRGLIEMKMELEPMNAADEA  
EQFVENTGEIFPGLMASGMAACAVHGGPRMGPIFGGMILSGKKIAQIIIEKLR

>Pyrobaculum aerophilum  
MELKIGRAIIRHAKLDDEYSDVDVAIVGAGPSGLAAKYLAEKGKLVVYERRFSFGGGIPGGNMLPKIVVQEEAVPILRDFKVRYKPAEDGLY  
VDPAAELIAKLAAGAVDAGAKIILGVHVDDVIFRGDPVRTGLLWIWTPIQMSGMHDPLYTQAKAVIDATGHDAEVVSVAARKVPELGIQVVGEKSA  
WSEVSEKLVVEHTGRVAPGLYVAGIAVCAYGLPRMGPIFGGMILMSGKKVAEVVKDLMMAEAHAVRA

>Pyrococcus abyssi (strain GE5 / Orsay)  
MLREVTISRAIESYYRDLLNLELDVAIVGAGPSGMVAAYYLAKGGAKVAIFEKKLSIGGGIWWGGMGFNKVVQEEAREIILDEFDIRYEEFEKGY  
YVADAIEVATTIASKTVKAGVKIFNMIEVEDLUVVKDNRVS GIVINWTPVLMTGLHVDPVTLTVEAKYVIDSTGHGAQVAQFLLKRLIERIPGEGAMWA  
EQGERLTVENTREVFPGLYVTGMAANAIAGAPRMGPIFGGMILSGKKIAQIELEKLNL

>Rikenella microfusus  
MEKLVSLGIVENYFEKLKNNLSDAAIVGGGPGSLVAAYYLAKAGRKVVLAYERKLAGGGMWGGAMMFNDIIVQQEALPILDELGVCYKPYREGACV  
VDSVHATSALVYAAATKAGATIFNCYSVEDVIFRDEAVAGLVNVAPVMREGMHVDPLMTAKTVLEGTGHDCMIARLVARNNVRNLTPTGEVAGER  
SLNVEQGERLTVENTKEIYPGLFVSGMAANGVSGSFRMGPIFGGMLMSGKAAELMIAKING

>Saccharolobus solfataricus (strain 98/2)  
MEVKIKQVDEVKISRYIIKETMEDWYQFVESDVIVGAGPSGLSAAYYLAKAGLKTFLVFERRLSFGGGIGGGAMLFHKLIIEKPADEILREVNIRLK  
EVEEGVVVDVDAEFLMAKLATAAIDAGAKIIHGVTDDVIFRENPLRVAGVAVEWTATQMASLHVDPFISAKAVVDATGHDAAEVISVAARKIPELGI  
VIPGEKSAYSERAAEELTVINTGKVAEGLYATGMATVEVKGLPRMGPIFGAMVLSGRKAAEELMI

>Smithella sp. SDB  
MLNETTISRRAILDAYFKKLDSCLELDVAIVGGGPGSLVAGYYLAKAGRKVALFERRLSIGGGIWGGGMMFNIAIVQEAQRQLLEFDLKGSEYAPGY  
YVLDADVDTATLIHKAVRAGLQVFNLNIAMEDVVIKNERVAGLVINWGAVIDTLKHVDPLTIHARYVIDGTGH PANVTEVLVRKMGRVLSNTPTGKMMG  
EKSMEAEGQELQTVENTREVYVPGLYVSGMAANAVFGGYRMGPVFGGMLLSGRKAAEELLAHL

>Spirochaetes bacterium ADurb.Bin215  
MLDETVISRAIETYMKKLTDLNSDVAIVGAGPSGFVAGYFLAKAGRKVVI FERALAVGGGMWGGGMGFNEIVVQEEGKAVLDEFDLPAVRYVQGY  
YTLDHSVRAVSALALRAVEAGVTVFNLGVVEDVVLHDERVSGLVLNWGTVMKAGIPVDPVTHSRCVLDTGHPAHVAEVLCRKMGVSLNTPTGKMMG  
EMMSDAEKGEKQTVENTREAYPGLFVSGMAANAVFGGYRMGPVFGGMLLSGRKAAEELLAHL

>Staphylothermus hellenicus (strain DSM 12710 / JCM 10830 / BK20S6-10-b1 / P8)  
MKFFPQONLYELSEGDSLTKLIDALYKKLSEIVKVDVAIVGAGPSGLTAAWKLGEKGYKVLVLERMLGVGGGMRGGSMLLPVGLIEDGEAAEIAREAG  
ARINKRNGLFVVDPSELAVRLASKAIAENGAIIPWGVLEDLITRGRGEDLTVVKGLINWTPPIYEAGWHVDPFYIEANAVV DATGHDGSLLRLVLA  
HPELKINIPGMSSQNVWIGEMVVEKTSMMVKGLFTGMSVAELYNTNRMGAIFGGMLVSGRKVADLIDDYFGKTRTLREQ

>Sulfolobales archaeon SCGC AB-777\_J03  
MASVRVPESKISR FIVEETMKDWDMDIVESDVIVGAGPSGMATAAYYLAKAGLKTFFRLGFGGGIGGGAMQFHRLVIEEPADEVLRFGVRLKK  
VDEGVYVVDAAEFLMAKLAASAIIDAGAKIILGVTDDVIFREDPPRVAVGVAWEWTATQMSGHLVDPFLFISAKAVV DATGHDAAEVISVAARKIPELNIS  
VPGEKSAYSEVAEQLVVNDTGPVAPGLYAAAGMAVCVEKSLPRMGPIFGAMVLSGRKVAELIIQDLRK

>Sulfolobus acidocaldarius  
MSDSIKIKAIDEVKISRYIIKQTMEDWMNFVENDVVIVGAGPAGMSAAYYLAKHGLKTFLVFERRLSFGGGIGGGAMLFHKLVIESPADEVILKEMNIR  
LEKVEDGVYIVDSDAEFLMAKLAASAIIDAGAKIIHGVTDDVIFRENPLRVAGVAVEWTATQ MAGHLVDPFISAKAVV DATGHDAAEVAVASRKIPEL  
GIVIPGERSAYSEMAEKLTVQTVGVPAGLYVAGMSVTEVRGLPRMGPIFGSMVLSGRKVAEDI IKDLRNS

>Synergistaceae bacterium  
MRLDEVTTIKAIMERYFDKFMMNLELDVAIVGGGPGSLVAGYFLAKAGRVALYERKLSVGGMWGGMLFNEIVVQEDAKRLLDELDVPTLPYKEA  
GYYTADSVEVTSTITSKAVKAGL VVFCNISVEDVVVKKDRISGLV INWTA VPMANLHVDP LSI RSRV IDATGHDTEVVAMVAKAPGRLLTPSKNI  
EGEKFMNPEEAERLTLNTKEVFPGLYVAGMACNATFGGPRMGPIFGGMMLSGEKVARLILKELSK

>Syntrophaceae bacterium PtaB.Bin038  
MALNEVTISRAIETYTKKLAHLDVDVAVVGGGPAGLVAAYFLAGAGRKVALYERKLSIGGGMWGGMMFNEIVVQAEAKGILDHFVVRTQEYAPG  
YYTADAIEAVTICCSRAIQAGAKVFNCITVEDVVIRDNRMGLVITWSPVEMTGLHVDP LTIHAKAVIDATGHDTEVLHVIER  
KADVTLNTPGKLMGERSMWSEKAERLTIDNTREICPGVYVAGMSANA AFGGPRMGPIFGGMMLSGRKVAEQILAGG

>Syntrophobacter sp. DG\_60  
MALDELIITQAIVERFSEKLKGCLMDVAIVGAGPSGLVAGYYLAKNGHRVAIFEKKLSIGGGMWGGMMFNQIVVQTEGKRILDEFI K TAPFSEG  
YYTADAIEAITTICSKACQAGVN FNTISVEDV L VREGRVIGL V INW STV EMAK L DV DPL TIR AQYVIEATGHATEVV KVIEKKMGE SLLPTGKII  
GEKSLWAEIGEADTIKNTKEA PGLFVCGMAANATFGSYRMGPVFGGMLLSGRKVAEQIL

>Syntrophobacteraceae bacterium  
MELNEITITQAIVDRFLEKFRNSLET DVAIVGGGPAGLVAAYFLAGAGRKVALYERKLSVGGMWGGMLFNEIVVQEEAKRLLDLFGVGSHHYRD  
YYTADSVEAISTLTSQAVKAGVMIFNCISVEDVV MRP ERVIGL V INW STV EMAK L DV DPL TIR AQYVIEATGHATEVV KVIEKKMGE SLLPTGKII  
GEKSMWSEVAEKLTLNTREVFPGLYVAGMAANATFGGPRMGPIFGGMMLSGEKVAQTLIDQ LKK

>Syntrophobacterales bacterium CG\_4\_8\_14\_3\_um\_filter\_49\_14  
MELNEITITKAIIERFSEKLI ACTEVDTAIVGGGPAGLVAAYFLAKVGKKVAIFEKKLSIGGGMWGGMMFNEIIIVQPEARELLDLFDVTRKYEAG  
YY SADAIEAVSTICSYATKAGARVFNCITVEDVMIREGRVIGL V INW TPV SMT GLHV DP L TIAKSTIDATGHATEVLRVIERKADVRLFTETGKLM  
GERSMWADRAERLTLENTREICPGVYVAGMSANA AFGGPRMGPIFGGMMLSGRKVAELIANG

>Syntrophus sp. (in: Bacteria)  
MELNEITITKAIIERFSEKLI ACTEVDTAIVGGGPAGLVAAYFLAKVGKKVAIFEKKLSIGGGMWGGMMFNEIIIVQPEARELLDLFDVTRKYEAG  
YYTADAIEAVSTICSYATKAGARVFNCITVEDVMIREGRVIGL V INW TPV SMT GLHV DP L TIAKSTIDATGHATEVLRVIERKADVRLFTETGKLM  
GERSMWADRAERLTLENTREICPGVYVAGMSANA AFGGPRMGPIFGGMMLSGRKVAELIANG

>Thermaaerovibrio acidaminovorans (strain ATCC 49978 / DSM 6589 / Su883)  
MELDERRISAVI VRRFMDRLLDMDLDVAIVGGGPAGLVA GHNLAREGFKVAMFERKLSLGGGMWGGMMFNQIVVQEEGAQVLREFGVRVLD EGEG  
YY SADSVEAVSTL ISSATRAGL RFNCVTAEDV TMREDRVVGLV ITWTPVEMAGL HVDP LTAIRSRF VIDATGHDINVVRVVERK VPGKLMTPTGRAE  
GEKSLWSHRAEELTLNTREVFPGLYVAGMSANATFGGPRMGPIFGGMMLSGRKAAQ L VS RALRGQGGRG

>Thermincola ferriacetica  
MHLDETVISRGIVQKYMEELMMDYMNTEVAIVGGGPGSMVAAYYLVKRGCKVALFDRKLAVGGGMWGGAMMFNKIVVQSAGKRILDEFAISCEEYERG  
YYVADAVESVTIATASMTVKAGCKIFNLIGAEDVMVEDGRVTGLVNWTPVQVNYYHVDPLVRAKYVIDGTGHPAEVTLTRKMGVRNLNTPTGGVA  
GEKPMNALKGEELDVVENTREVFPGLYVTGMAANAAFGSHRMGPVFGMLLSGEKAAMEIAARLGK

>Thermococcales archaeon 44\_46  
MLRDVTISRAIIETYFKELLEHNLDAIVGAGPSGMVAAYYLAKGGAKVAIFEKLSIGGGIWGGGMGFNKIVVVEEAEKIIEFGVRHEEFEEGY  
YVADAIEVATTIASKNIKAGAKIFNMVEVEDLVLVKENRVAGIVINWTPVKMTDLHVDPLTVEAKFVIDSTGHGAQVTQLLERKGLIERVPGESAMWA  
EMGEKLTVEHTKEIYPGLYVTGMAANAVAGAPRMGPIFGGMFLSGRKAAFEILEKLKK

>Thermococcus celer Vu 13 = JCM 8558  
MLKDVEVSRAIEAYTKDILDLSKLKDVAIVVGAGPSGMVAAYYLARGGAKVAIFEKLSVGGSIGGGIWGGAMGFNRIVVEESAREILDEFGVDYEEFKPGL  
YVADAIEVATTMASKTVKAGVKVFNMVEVEDLVLVKGDRVAGVVNWTPVKMTGLHVDPLTVEAEFVIDSTGHGAQITGHLLKRLIEELPGECPMWA  
EMERLTVEHTKEVFPGLYVTGMAANAVAGAPRMGPIFGGMFLSGRKAAFDILERLG

>Thermodesulfatator autotrophicus  
MALEDEIKISRAIITYFKKLTDYLEMDDAVVGAGPSGLMAAYKLASEGFKVAFVFFRSLSIGGGIWGGMMFNEIVVQEEGARLLKEIGVRTEPWNGG  
EYYTADAVEVACILAASKVQAGAKIFNLIMVEDVMVRDNRVVGVLVNWNSATEIAGLHVDPЛАVKAKYVVEATGHETAVLQVMQKLGAKLNTEGKV  
MGEKSMWAEVANLTVDTYDREVYPGVFVAGMAANATFGAYRMGPIFGGMLLSGERAQLIAERLRQ

>Thermodesulfobacterium geofontis (strain OPF15)  
MELKIQRAIVKFGMEDLYEYSDVDLIVGAGPSGLTSAYKLADKGFKLVYEKRLSFGGGIGGGGNMIPKIVVQEEALPILKDFKIKYKEAEKNLYT  
IDPAELIAKLAvgALDAGAKIILGVHVEDDVIRDNPPRTGVLWRWTIAIEISGLHVDPLOYTQSKALIDATGHGAEIVQIAAEKNPELNIIKGEKSN  
WSEVSEKLVVDYTGKVAEGLYVTGIAVCEVFGLPRMGPIFGGMILMSGKKIAEIIIEKDLRG

>Thermodesulfobium acidiphilum  
MTKHFLNPVTDNVSKLILKHYFESITDALTSVIIIVGGPSGLTAARELGNSGYKVVIMERKLSPGGTWGGSMSFNKVIQKDLKDYLNELEIPF  
VEDLDALVVDSCLFASQLIAKALKTONVLFNLMVTDLEYTNNAITGVVNNNTGIETAGLHVDPMVFTQKAVALDATGHDAIAANIYSKRVQLPLRK  
EHFMNAVQGEEDTVNNNTKMLANGLFVSGMAANNVDDGSSRMGPIFGGMILMSGKKIAAKLIMEYIKTV

>Thermodesulfobacter norvegica  
MAELNEIIITRAIIDRYHAKITGNLDVDAIVGAGPSGLVAGYHLAQKGYRTIFERKLSVGGGIWGGMLFNEIVVQDEARRILEEFGVRVNRYEE  
NYYTADAIETVSLLAARAIQKGVTILNGITVEDVVMRPNRIVGLVILWSAVEIAGLHVDPЛАIRAKYIVDATGHDTENVKVVHKVPGRMLPTGNI  
EGEKSWMSEEAKLTLENTREVFPGLFVAGMAANATFGPPRMGPIFGGMLLSGEKVAHLIDERLKQGS

>Thermoplasma archaeon HGW-Thermoplasma-1  
MTVIDEVVTRKIFDRFSREFLDHLDVDVALVGGGPANMVAAYHLAKAGKKVVLFERKLAPGGMWGGGMNFVIVIQEALPIMNEFGIKVEGSND  
GYYTADSVECVAKKLAKSIDSGARVNPNSMTVQDMVIRDDDNKVGVVIWVGDITKHMVDPITRAKFVVDGTGHPCEVNVAKKAGRRLRTPSGK  
VEGERSMWAEVGEETTVNNNTVEIYPGLYVAGMAANVGMAPRMGPIFGGMILSGKKVAEMILQKLGA

>Thermoprotei archaeon ex4572\_64  
MSIVRKVSEDEISKAIINEALKESLVDVDAVVVGSPGSLTCSYYLAKYGLKTVLIERRLSFGGGIGGGMLLPSIAIESPAEELIHDEFVLNIK  
KVRDGLYVVMNPAAFIKLAASKAINAGVKVLLGVSEDVIFRSNPLRIAGVVIWSAHISQLWDPLFIKAKAVIDATGHDAEVVNIVSKIPDFKL  
AIKGEKSACSIEADELIISYSGKVEGLYVTGMATAKVYGLPRMGPIFGGMVLSGKKTAEEVYRDLSELNR

>Thermoproteus uzonensis (strain 768-20)  
MRKINLIWKLRGAMELKIGRAIIRHGAEDLYEYSDVDAIVGAGPSGLTAARYLAEGKLKVIVLERRFSFGGGIGPGGNMYPKIIIVQEEALPILRDF  
KVRYKPGADGLYAVDPAELIAKLAAGAIDAGAKILLGVHVDVIFRGPRTGGLWIWTPIQMSGMHVDPLYIQTКАVVDATGHDAEVVSVAARKV  
PELGIQLQGEKSAWSEVSEKLVVEHTGKVAAPGLYVAGMAAAVFGLPRMGPIFGGMILMSGKKVAEIVAKDLAAEVHAV

>Thermosiphon africanus  
MWDEVSKIIIVERFFEKLNNDLNVDVAIVGGGPGSALSASYYLSKKGLKVAIFEAKEPNEPGGGTWGGMMFNELVVENDIKSFLDELGMNYLIKDNFIS  
VDSVHFASSLLYNTKAGAVLFNNVIVEEDIAFYENKVNNGIVINWAPVIRQKLHVDPITIMAKFVVDGTGH PANVNMVLDRGIDIDLPIKGIREYPM  
NAKEGEKFVVENTKEVFPGLYVGMMAAVSVGGGPRMGPIFGGMILSGKKVAEIKLVEVSK

>Thermosiphon melanesiensis (strain DSM 12029 / CIP 104789 / BI429)  
MWDELEISKIIIVNGFFEKFNALDWDVAIVGGGPGSALTASYFLTKNGFKVVIFEKNDPGGGTWGGGMFLNELVVVEELEWMLKEFGMNYKRLNGFIS  
IDSVHFASSLLYNTKVGTKIFNNNIVEDILMNEENRLCGVVINWAPVIKQRLHVDPITVKAKYVVDGTGH PASVVQMIIDRNLEVELPLDKIREFPM  
NAKEGENFVLKNTKEVFPGLFVGMMAAVSVGGGPRMGPIFGGMILSGKKVAEIKLVEVSK

>Thermosulfidibacter takaii (strain DSM 17441 / JCM 13301 / NBRC 103674 / ABI70S6)  
MLDEKIITKAIIESYTQNLLDYIDMDVAIVGAGPAGLTCAYYLAKEGFKVGFERKLSIGGGIWGGAMFNEIVLQEEALPIVQEMEVSYKPYKEKG  
YYVINAVERFACALGLKAIRAGAKIFNLWSAIDVKVKGEDERVNGLWLLWTPVDTAGLHVDPITVEAKYVVDGTGHDAEIANVVKKLKKLATPTGD  
VAGERPMWAEGEKEATEFTGEVYVGLFVIGMAAVACYGKHRMGPIFGGMILSGKKAAKMILECLK

>Thermosulfurimonas dismutans  
MALDEVKITQAIVERFTEKLEALELDVAVVGAGPSGLMAAYKLAKEGFKVAIFERKLSIGGGIWGGMMFNEIVVQEEGARLLKEIGVEARFWQED  
YYTADSVETVCAVGLYAAKAGAMIFNLISVEDVMVRKDRVGLVINVWATVEMGGHLHVDPLAIRSKYVVESTGHELSVLHIMQKLGVLMTPSGKIE  
GEKSLWADVAETTLENTREVFPGVFVAGMAANATFGSYRMGPIFGGMILSGEKVAQEIAARLK

>Thermosyntropha lipolytica DSM 11003  
MVINDIKITRSIIIEYYYAFTRDFLDCDVIVGGGPAGMTAAAYTAQQGLRTVVLESRLSPGGMWGGGMFFNQIVFQPEAGEILQELGISYTANREG  
YLVVP SYRAVASLILAADRAGARILNGITAEDIMRENRCGVVINWTAAVKLGMDPLCIGGKVVIDATGHDAGIVRTYLDKGSSLPEDEEERI  
RTSSMWA AKGEEMVVEYTRFITEGLIACGMSVSSLFNTPRMGP IFGGMLFSGRKAAELALDYIRKVKA

>Thermovirga lienii (strain ATCC BAA-1197 / DSM 17291 / Cas60314)  
MKLDEKIISKAITTRYQKILSHIQVDVAIVGGGPSGLVAGYYLAKEHRVALFERKLISVGGMWGGMLFNEIVVQEDAKEILEDGFVRVQFWEDA  
GYYTADAIIESVCISITSKAIQAGLTVFNCISVEDVSVEGDRITGLVINWTPVEMSGLHVDP LSI GASFVI DATGDTEVVHMVAKAPGKLMTPSGDI  
EGEKFMCPDEAEKKVENTKEVFPGLYVAGMACNATFGGPRMGP IFGGMLLSGRKVAALISQRLK

>Treponema sp. CETP13  
MLEYNVSKGILD SYHTKLKSALDS DAIIVGSGPSGLVAGYFLAKAGKKVVMFERELAPGGGIWGGMFNDVMQEEAATILSEIGVELPEVKDNFY  
TIDSVYLASTLISKAVEAGVTLLNMISIEDII FAKDESIGGVVLNWPVKEHMHVDP LMAISRCVLDATGH PSEIVNLTRKNEITLNTKTGKVMG  
ERSLKCKK AELATAENTCEIYPRLFVSGMAANGVAGAYRMGPVFGGMIRSGKKVAEQLQCIDTEAPIYD

>Vulcanisaeta distributa (strain DSM 14429 / JCM 11212 / NBRC 100878 / IC-017)  
MAGIYI SESSITRAIMRSALKMLDEYSSVDVAIVGAGPSGM TAAYYLA KAGLKTIVLERRSF GGGIGGAASHLPSIVVEYPASDILSKDFGVRLQD  
MGDGLFAVDPAEMIAKLA VRAIDAGAKFLLG VHDDVI RDNPPRVA GLAVY WSTVQ MAGVHTDPFFIEAKAVV DATGHDAEV AAVTTRKNPDGLA  
IHGEKSAHASVAE DLVVKY TGRVMEGLYVTGMAVA AVYGLPRMGP IFGSMIMSGKRV AEIINDLRR

>Zestosphaera tikiterensis  
MEPLEAKISKIWIWKETLNDWLKLSNV DVVVVGAGPSGMV TAKYLADSGIKTLVLERRLSFGGGIGGGMLMHKVVVDSKALNI LDDFKIKYSRS DYE  
GLYVVDASELMAKLAAGAIDSGAKIVNGITVEDLIVRDNPFRVEGVVIQWSAVNLSGLHVDP LFI YSKAVV DATGHDAEV LKVLSRKNPEVNLKIPG  
EKSAYAELSEELVVVKHSGKVLPGLYVSGMAVA ALYGIYRMGPIFTGMLLSGKKVAEEIAKDLRGGSQ

**Supplementary Table 4 Occurrence of ROS defense genes in genomes of prokaryotes whose non-Cys THI4s have crystal structures**

The predicted proteomes in NCBI were searched by BlastP with the indicated query sequences. Significant hits (e-value < 1e-05) are indicated with a plus sign.

| Enzyme                  | Query <sup>a</sup>           | Organism                         |                             |                                 |   |
|-------------------------|------------------------------|----------------------------------|-----------------------------|---------------------------------|---|
|                         |                              | <i>Thermovibrio ammonificans</i> | <i>Methanococcus igneus</i> | <i>Methanococcus jannaschii</i> | <i>Methanothermococcus thermolithotrophicus</i> |
| Catalase-peroxidase     | WP_013536945                 | +                                | -                           | -                               | -   |
| Cytochrome c peroxidase | WP_013537906                 | +                                | -                           | -                               | -   |
| Cytochrome bd complex   | WP_013537252<br>WP_013537253 | +                                | -                           | -                               | -   |
| Heme-catalase           | WP_000077872                 | -                                | -                           | -                               | -   |
| Mn-catalase             | WP_000488336                 | -                                | -                           | -                               | -   |

<sup>a</sup> GenBank identifier.