Novel Transcriptional Regulons for Autotrophic Cycle Genes in *Crenarchaeota*

Semen A. Leyn,^a Irina A. Rodionova,^b Xiaoqing Li,^b Dmitry A. Rodionov^{a,b}

A. A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia^a; Sanford-Burnham Medical Research Institute, La Jolla, California, USA^b

ABSTRACT

Autotrophic microorganisms are able to utilize carbon dioxide as their only carbon source, or, alternatively, many of them can grow heterotrophically on organics. Different variants of autotrophic pathways have been identified in various lineages of the phylum *Crenarchaeota*. Aerobic members of the order *Sulfolobales* utilize the hydroxypropionate-hydroxybutyrate cycle (HHC) to fix inorganic carbon, whereas anaerobic *Thermoproteales* use the dicarboxylate-hydroxybutyrate cycle (DHC). Knowledge of transcriptional regulation of autotrophic pathways in *Archaea* is limited. We applied a comparative genomics approach to predict novel autotrophic regulons in the *Crenarchaeota*. We report identification of two novel DNA motifs associated with the autotrophic pathway genes in the *Sulfolobales* (HHC box) and *Thermoproteales* (DHC box). Based on genome context evidence, the HHC box regulon was attributed to a novel transcription factor from the TrmB family named HhcR. Orthologs of HhcR are present in all *Sulfolobales* genomes but were not found in other lineages. A predicted HHC box regulatory motif was confirmed by *in vitro* binding assays with the recombinant HhcR protein from *Metallosphaera yellowstonensis*. For the DHC box regulon, we assigned a different potential regulator, named DhcR, which is restricted to the order *Thermoproteales*. DhcR in *Thermoproteus neutrophilus* (Tneu_0751) was previously identified as a DNA-binding protein with high affinity for the promoter regions of two autotrophic operons. The global HhcR and DhcR regulons reconstructed by comparative genomics were reconciled with available omics data in *Metallosphaera* and *Thermoproteus* spp. The identified regulons constitute two novel mechanisms for transcriptional control of autotrophic pathways in the *Crenarchaeota*.

IMPORTANCE

Little is known about transcriptional regulation of carbon dioxide fixation pathways in *Archaea*. We previously applied the comparative genomics approach for reconstruction of DtxR family regulons in diverse lineages of *Archaea*. Here, we utilize similar computational approaches to identify novel regulatory motifs for genes that are autotrophically induced in microorganisms from two lineages of *Crenarchaeota* and to reconstruct the respective regulons. The predicted novel regulons in archaeal genomes control the majority of autotrophic pathway genes and also other carbon and energy metabolism genes. The HhcR regulon was experimentally validated by DNA-binding assays in *Metallosphaera* spp. Novel regulons described for the first time in this work provide a basis for understanding the mechanisms of transcriptional regulation of autotrophic pathways in *Archaea*.

The *Thermoproteales* and *Sulfolobales* are two groups of thermophilic archaea from the phylum *Crenarchaeota*. Members of both groups can grow under autotrophic conditions by using carbon dioxide as a carbon source (1, 2). Studies of model organisms revealed the presence of two distinct autotrophic carbon dioxide fixation pathways in these crenarchaeal lineages (3–12). *Thermoproteus neutrophilus* and other *Thermoproteales* possess the dicarboxylate/4-hydroxybutyrate cycle (DHC), whereas *Metallosphaera sedula* and related *Sulfolobales* use the 3-hydroxypropionate/4-hydroxybutyrate cycle (HHC). These two autotrophic pathways share seven common reactions of conversion of succinyl coenzyme A (succinyl-CoA) to acetyl-CoA; however, other steps of incorporation of carbon dioxide into organic compounds differ between the cycles (Fig. 1). In the *Sulfolobales*, two molecules of carbon dioxide in the bicarbonate form are incorporated into metabolic intermediates by the bifunctional enzyme acetyl-CoA/propionyl-CoA carboxylase encoded by the *accABC* genes (13). In the *Thermoproteales*, one molecule of carbon dioxide is used to convert acetyl-CoA to pyruvate by one of two pyruvate synthases encoded by *porABDG* and *korAB*, respectively, and one additional molecule of bicarbonate is used to convert phosphoenolpyruvate to oxaloacetate by phosphoenolpyruvate carboxylase (*ppcA*).

Both pathways produce important intermediates of the central carbon metabolism, including acetyl-CoA, pyruvate, phosphoenolpyruvate, and succinyl-CoA (1).

Interconnection of the DHC and HHC pathways with the central carbon metabolism requires transcriptional regulation of autotrophic carbon dioxide cycle genes. Indeed, microarray transcriptional response analysis of *M. sedula* showed significant upregulation of the characteristic HHC enzymes under au-

Received 30 March 2015 Accepted 29 April 2015

Accepted manuscript posted online 4 May 2015

Citation Leyn SA, Rodionova IA, Li X, Rodionov DA. 2015. Novel transcriptional regulons for autotrophic cycle genes in *Crenarchaeota*. *J Bacteriol* 197:2383–2391. doi:10.1128/JB.00249-15.

Editor: I. B. Zhulin

Address correspondence to Dmitry A. Rodionov, rodionov@burnham.org.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/JB.00249-15>.

Copyright © 2015, American Society for Microbiology. All Rights Reserved. doi:10.1128/JB.00249-15

FIG 1 Pathways of autotrophic carbon dioxide fixation in *Crenarchaeota*. Genes from the HhcR and DhcR regulons are in pink and blue, respectively. The numbers of genomes in which HhcR or DhcR sites were observed in regulatory gene regions are indicated by numbers in pink and blue circles. Enzyme reactions are indicated by italic numbers next to arrows: 1, succinyl-CoA reductase; 2, succinic semialdehyde reductase; 3, 4-hydroxybutyryl-CoA synthetase; 4, 4-hydroxybutyryl-CoA dehydratase; 5, crotonyl-CoA hydratase; 6, 3-hydroxybutyryl-CoA dehydratase; 7, acetoacetyl-CoA β -ketothiolase; 8, pyruvate synthase; 9, pyruvate-water dikinase; 10, phosphoenolpyruvate carboxylase; 11, malate dehydrogenase; 12, fumarate hydratase; 13, fumarate reductase; 14, succinyl-CoA synthetase; 8', acetyl-CoA carboxylase; 9', malonyl-CoA reductase; 10', malonate semialdehyde reductase; 11', 3-hydroxypropionyl-CoA synthetase; 12', 3-hydroxypropionyl-CoA dehydratase; 13', acryloyl-CoA reductase; 14', propionyl-CoA carboxylase; 15', methylmalonyl-CoA epimerase; 16', methylmalonyl-CoA mutase.

totrophic versus heterotrophic conditions (7, 14). However, the mechanism of transcriptional regulation of the HHC pathway in the *Sulfolobales* is currently unknown. In *T. neutrophilus*, the combined analysis of enzyme activity, the proteome, and transcription revealed that most of the DHC enzymes were induced in cultures grown autotrophically on carbon dioxide in comparison with cells grown in the presence of organic substrates, such as acetate (15). Genes encoding four key enzymes from the DHC pathway in *T. neutrophilus* are organized into a single cluster that contains two divergently oriented leaderless transcripts, *scr-4hbl* and *4hbd-frdA-frdB*, and both of these transcripts were significantly down-regulated in cells grown in the presence of acetate. A hypothetical protein encoded by the *Tneu_0751* gene was found to bind with high affinity to the 127-bp promoter region of the *scr-4hbl* and *4hbd-frdA-frdB* operons (15). Because of the constitutive expression of *Tneu_0751*, it was proposed that this potential autotrophic regulator responds to some inductor or effector molecule that governs its promoter binding. However, other target genes and

DNA-binding sites of the putative regulator Tneu_0751 were still unknown.

We previously applied the comparative genomics approach for reconstruction of DtxR family regulons in diverse lineages of *Archaea* (16). In this work, we used this bioinformatics approach to infer novel *cis*-regulatory elements and to reconstruct regulons for autotrophic pathway genes in the *Thermoproteales* and *Sulfolobales* lineages of *Archaea*. We report the identification of two distinct DNA motifs that potentially control the DHC and HHC genes, respectively, in these two lineages. The novel DNA motif in *Thermoproteales* is a predicted target of Tneu_0751 and its orthologs (named DhcR), whereas the *Sulfolobales*-specific DNA motif was predicted to be associated with a novel TrmB-like transcriptional regulator (named HhcR). DNA-binding assays with the purified HhcR protein from *Metallosphaera yellowstonensis* confirmed its predicted DNA motif. A comparison with global transcriptomic and proteomic data previously obtained for *M. sedula*, *Metallosphaera cuprina*, and *T. neutrophilus* grown either

autotrophically or heterotrophically provided additional validation and enrichment of the genomic reconstruction of HhcR and DhcR regulons.

MATERIALS AND METHODS

Bioinformatics tools and databases. Complete genomes of *Crenarchaeota* were downloaded from GenBank (17). The order *Thermoproteales* includes seven species from the genus *Pyrobaculum* (*Pyrobaculum arsenaticum* DSM 13514, *Pyrobaculum islandicum* DSM 4184, *Pyrobaculum aerophilum* strain IM2, *Pyrobaculum neutrophilus* V24Sta [also known as *Thermoproteus neutrophilus* V24Sta], *Pyrobaculum. oguniense* TE7, and *Pyrobaculum* sp. strain 1860) and two *Thermoproteus* species (*Thermoproteus tenax* Kra 1 and *Thermoproteus uzoniensis* 768-20). The order *Sulfolobales* includes four representative species from the genus *Sulfolobus* (*Sulfolobus islandicus* Y.N.15.51, *Sulfolobus acidocaldarius* DSM 639, *Sulfolobus solfataricus* P2, and *Sulfolobus tokodaii* strain 7), two *Metallosphaera* species (*M. cuprina* Ar-4 and *M. sedula* DSM 5348), and *Acidianus hospitalis* W1. In addition to these complete genomes, we analyzed draft genomes of two related species recently isolated from Yellowstone National Park (YNP), namely, *M. yellowstonensis* MK1 (8) and *Pyrobaculum yellowstonensis* WP30 (18), which were downloaded from the Integrated Microbial Genomes database (19). Identification of orthologs was performed using BLAST against the “nr” database (20) and the bidirectional best-hit method implemented in the Genome Explorer software (21). Multiple alignments were built by MUSCLE (22). Identification of a common DNA motif in gene upstream regions was performed by MEME (23). Motif-specific positional weight matrices were built using the SignalX program (24). For functional protein annotations, distant homology to characterized proteins was determined using BLAST against the Swiss-Prot database. Autotrophic pathway enzymes were annotated using a compilation of recent publications (3–9). Protein domain families were determined by sequence similarity search tools implemented in the Pfam database (25). Conserved gene neighborhoods were analyzed using the Gene tree tool in the MicrobesOnline Web server (26). Logo diagrams for DNA motifs were created using the WebLogo program (27).

Regulon identification and analysis. For reconstruction of autotrophic regulons in archaeal genomes, we used the comparative genomics approach that was previously extensively used for the inference of *cis*-acting regulatory elements in regulatory networks of *Bacteria*. Regulons are sets of genes that are coregulated by the same transcription factor (TF) in a single genome. The comparative genomics approach is based on the assumption that regulons have a tendency to be conserved between related genomes containing orthologous TFs (28). Simultaneous analysis of multiple genomes from the same taxonomic group allows one to make reliable predictions of TF-binding sites (TFBSs) even with weak recognition rules. After validation of TFBSs by comparative genomics, the improved positional weight matrix is used to scan the genomes for the presence of additional TFBSs contributing to regulon expansion beyond the initial training sets. This leads to another cycle of iterative refinement aimed at maximizing the coverage and consistency of the reconstructed regulons.

We first collected training sets of upstream DNA regions of potentially coregulated genes involved in the autotrophic pathways and their orthologs in the two groups of genomes analyzed. The *Sulfolobales*-specific training set included the HHC pathway genes that were significantly up-regulated under autotrophic conditions in the transcriptomic analyses of *M. sedula* (7, 14) (see Table S1 in the supplemental material). For the *Thermoproteales* group, the training set included the DHC pathway genes from the comparative proteome and enzyme activity studies of autotrophically and heterotrophically grown cells of *T. neutrophilus* (15), including *Tneu_0418* to *Tneu_0422*, *Tneu_1797* to *Tneu_1795*, and *Tneu_1204*. For each training set of genes, we determined orthologs in related genomes, analyzed the potential operon structure of their loci, and collected their upstream DNA regions. The prepared training sets of upstream regions from the *Sulfolobales* and *Thermoproteales* lineages were

used for a search of lineage-specific conserved DNA motifs using MEME. We allowed MEME zero or one site per upstream sequence. The identified putative regulatory DNA motifs were refined using the phylogenetic footprinting approach (28) by construction of multiple alignments of orthologous upstream regions (see Fig. S1 in the supplemental material). The validated candidate regulatory sites were used to build a positional weight matrix (or profile) using the SignalX software. Each genome in both archaeal lineages was scanned with the constructed profile using Genome Explorer software (21), and genes with candidate regulatory sites in their upstream regions were selected. We analyzed only 5' untranslated gene regions up to 250 nucleotides upstream of the translation start site. *z* scores of candidate sites were calculated as the sum of the respective positional nucleotide weights. The threshold for the site search was defined as 10% lower than the weight of the weakest site in each of the training sets. All identified candidate regulon members were verified by the consistency check approach to eliminate false-positive site prediction (28). This approach is based on the assumption that regulatory interactions tend to be conserved in closely related species. The reconstructed regulons included groups of orthologous genes that have potential regulatory sites in three or more genomes.

Gene cloning and protein purification. The *hhcR* gene (GenBank accession no. EHP68520.1; locus tag MetMK1_29590) from *M. yellowstonensis* MK1 was synthesized by GenScript Inc. with optimized codons for expression in *Escherichia coli*. The synthesized gene fragment was cloned into the pET-49b expression vector (Novagen Inc.), and the HhcR protein was expressed with a C-terminal thrombin-His tag in *E. coli* BL21(DE3) under the T7 promoter. Cells were grown in lysogeny broth medium (50 ml), induced by addition of 0.2 mM isopropyl- β -D-thiogalactopyranoside, and harvested after 4 h of additional shaking at 37°C. The harvested cells were resuspended in 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid buffer (pH 7) containing 100 mM NaCl, 0.03% Brij 35, 2 mM β -mercaptoethanol, and 2 mM phenylmethylsulfonyl fluoride (Sigma-Aldrich). The cells were lysed by incubation with lysozyme (1 mg/ml) for 30 min, followed by a freeze-thaw cycle and sonication. Recombinant HhcR protein was purified to homogeneity using Ni²⁺ chelation chromatography. First, for purification of the soluble fraction after centrifugation, Tris-HCl buffer (pH 8) was added to the supernatant (50 mM final concentration), which was loaded onto a Ni-nitrilotriacetic acid-agarose minicolumn (0.3 ml) from Qiagen Inc. After washing with starting buffer containing 1 M NaCl and 0.3% Brij 35, the bound proteins were eluted with 0.3 ml of the same buffer supplemented with 250 mM imidazole. Second, the insoluble protein fraction was refolded by resuspension in 8 M urea buffer containing 100 mM Tris (pH 7), 1 M NaCl, 0.3% Brij 35, and β -mercaptoethanol with centrifugation, and the supernatant buffer was exchanged with 2 M urea-At buffer by dialysis. After refolding, the protein fraction was loaded onto a Ni-nitrilotriacetic acid-agarose minicolumn and purified to homogeneity by the same procedure as for the soluble protein fraction. The purified proteins were electrophoresed on a 12% (wt/vol) sodium dodecyl sulfate-polyacrylamide gel to monitor size and purity (>90%). The protein concentration was determined with a Quick Start Bradford Protein Assay kit from Bio-Rad. The HhcR monomer with a C-terminal thrombin-His tag has a predicted molecular mass of 13.8 kDa. The molecular mass of the purified recombinant HhcR protein was calculated by gel filtration to be ~24 kDa, which corresponds to the dimer.

DNA-binding assays. The interaction of the purified recombinant *M. yellowstonensis* HhcR protein with its cognate DNA-binding sites in *M. yellowstonensis* was assessed using two techniques: electrophoretic mobility shift assay (EMSA) and fluorescence polarization assay (FPA). The single-stranded labeled and unlabeled DNA oligonucleotides were synthesized by Integrated DNA Technologies (see Table S5 in the supplemental material). The double-stranded DNA (dsDNA) fragments were obtained by annealing synthesized complementary oligonucleotides at a 1:10 ratio of 5' labeled with 6-carboxyfluorescein (for FPA) or biotin (for EMSA) to unlabeled complementary oligonucleotides. Using the FPA as-

FIG 2 Predicted DNA motifs associated with autotrophic genes in *Crenarchaeota*. (A) A-box (HhcR) motif in the *Sulfolobales*. (B) DhcR motif in the *Thermoproteales*.

say, we tested three 28-bp DNA fragments and one 37-bp fragment containing the predicted HHC boxes. The 6-carboxyfluorescein-labeled ds-DNA fragments (10 nM) were incubated at 37°C with increasing concentrations of the thrombin-His₆-tagged HhcR protein (200 to 1,000 nM) purified from the soluble fraction in a 100-μl reaction mixture in 96-well black plates (VWR, Radnor, PA) for 30 min. The binding buffer contained 50 mM phosphate buffer (pH 6), 50 mM KCl, 5% glycerol, 0.5 mM EDTA, 2 mM dithiothreitol. Herring sperm DNA (1 μg) was added to the reaction mixture as a nonspecific competitor DNA to suppress nonspecific binding. The fluorescence-labeled DNA was detected with the FLA-5100 fluorescence image analyzer. Using EMSA, we tested the *hhcR* (MetMK1_29590) DNA fragment containing the predicted HHC box. The biotin-labeled 46-bp DNA fragment (0.25 nM) was incubated with increasing concentrations of the recombinant HhcR protein (7.5 to 375 nM) refolded from the insoluble fraction in a total volume of 20 μl. The binding buffer contained 50 mM Tris-HCl (pH 8.0), 0.2 M NaCl, 5 mM MgCl₂, 1 mM dithiothreitol, 0.05% NP-40, 2.5% glycerol, and 1 μg herring sperm DNA. After 25 min of incubation at 37°C, the reaction mixtures were separated by electrophoresis on a 5% native polyacrylamide gel (100 min; 90 V; room temperature). The DNA was transferred by electrophoresis onto a Hybond-N⁺ membrane and fixed by UV cross-linking. The biotin-labeled DNA was detected with the LightShift chemiluminescent EMSA kit (Thermo Fisher Scientific Inc., Rockford, IL, USA). Additional DNA fragments of the *hoxN* and *thi4* gene upstream regions (see Table S5 in the supplemental material) were used as a negative control.

RESULTS AND DISCUSSION

Identification of the HhcR regulon in *Sulfolobales*. A novel putative regulon for the HHC pathway in the *Sulfolobales* genomes was initially identified by analysis of gene expression data in *M. sedula* (7, 14). Using microarray data for autotrophic and heterotrophic growth conditions, we selected the most upregulated HHC pathway genes, namely, *accAB*, *hbcD*, *hbcS3*, and *mcmB* (see Table S1 in the supplemental material). By applying the motif recognition procedure to a training set of upstream regions of these genes in *M. sedula* and orthologous genes in other *Sulfolobales*, we found a common 18-bp DNA motif named the HHC box (Fig. 2A). The consensus sequence for this tandem-repeat motif is WWTWWAAAC, where “W” denotes A or T. A recognition profile was constructed and used to scan the *Sulfolobales* genomes, resulting in identification of additional candidate sites upstream of genes from the HHC and other pathways (see Table S2 in the supplemental material).

The identified HHC box sites upstream of the HHC genes are mostly conserved across the studied *Sulfolobales* genomes and of-

ten overlap a putative archaeal promoter consisting of a TATA box and a purine-rich B recognition element (see Fig. S1A in the supplemental material). Other taxonomic lineages of *Archaea* lack similar HHC box sites upstream of the autotrophic pathway genes, suggesting that the HHC box regulon (termed the HhcR regulon) is restricted to the order *Sulfolobales*. The analysis of the genome context and the reconstruction of the HhcR regulons in the *Sulfolobales* genomes are outlined below and summarized in Table 1.

The reconstructed HhcR regulons include genes encoding enzymes for 13 out of 16 reactions from the HHC pathway in *Sulfolobales* (Fig. 1). Genes encoding acryloyl-CoA reductase (*acr*), malonate semialdehyde reductase (*msr*), succinate semialdehyde reductase (*ssr*), 3-hydroxypropionyl-CoA synthetase (*hpcS*), 4-hydroxybutyryl-CoA dehydratase (4hbd) (*hbcD*), 4-hydroxybutyryl-CoA synthetase (*hbcS*), methylmalonyl-CoA mutase (*mcmA* and *mccB*), methylmalonyl-CoA epimerase (*mce*), acetyl-CoA/propionyl-CoA carboxylase (*accAB* and *accC*), and acetoacetyl-CoA-ketothiolase (*ack*) have predicted HHC box sites that are conserved in the majority of the studied genomes of *Sulfolobales*. In contrast, the malonyl-CoA/succinyl-CoA reductase gene *mcr* has the predicted HHC box sites in only three *Sulfolobus* spp. (Table 1). The previously established bifunctional crotonyl-CoA hydratase/3-hydroxybutyryl-CoA dehydrogenase (*hpcD-hbd*) from the HHC pathway, which is encoded by *Msed_0399* in *M. sedula* (6), does not belong to the reconstructed HhcR regulon in the *Sulfolobales* spp. (see Table S1 in the supplemental material). In contrast, the HhcR regulons in *Metallosphaera* spp. and *Acidianus hospitalis* contain another candidate 3-hydroxybutyryl-CoA dehydrogenase gene, *hbd2* (e.g., *Msed_1423* in *M. sedula*), which is 46% identical to *Msed_0399*.

Interestingly, the *Sulfolobales* have at least three 4-hydroxybutyryl-CoA synthetase gene candidates. *Msed_0406* (*hbcS1*) and *Msed_0394* (*hbcS2*), which were recently identified by transcriptomics in *M. sedula* and showed weak catalytic efficiency with 4-hydroxybutyrate *in vitro* (7), are moderately conserved members of the HhcR regulon in *Sulfolobales* (see Table S2 in the supplemental material). HbcS1 has orthologs only in *Metallosphaera* spp., and it is preceded by an HhcR site only in *M. yellowstonensis*. HbcS2 has orthologs in all *Metallosphaera* spp. and three *Sulfolobus* spp. and in *Acidianus*, and it is predicted to be coregulated by HhcR in three out of six genomes, excluding *M. sedula*. Thus, HbcS1 and HbcS2 can be potentially involved in the HHC pathway, but not in all *Sulfolobales*. The third previously identified candidate for 4-hydroxybutyryl-CoA synthetase, *Msed_1422* (*hbcS3*), which was inactive *in vitro*, potentially because it was produced in an inactive form (9), has orthologs in all but one *Sulfolobales* genome (the exception is *S. acidocaldarius*). All *hbcS3* orthologs in *Sulfolobales* are preceded by high-scored HhcR sites, indicating that it is a strong candidate for involvement in the HHC pathway.

In addition to genes from the HHC pathway, the reconstructed HhcR regulons in the *Sulfolobales* genomes contain genes involved in carbon and energy metabolism, including phosphoglycerate mutase (*pgm*), propionyl-CoA/acetyl-CoA synthetase (*acsA*) (9), pyruvate synthase (*porGDAB*), lactaldehyde dehydrogenase (*ladH*), thiosulfate::quinone oxidoreductase (*doxDA*), and ferredoxin-NADP reductase (*fnr*), as well as a number of other genes of yet-unknown function (Table 1). For instance, two predicted members of the HhcR regulons encode putative enzymes from the

TABLE 1 Distribution of predicted HhcR regulon members across the *Sulfolobales* genomes

Gene name	Example locus tag	Presence ^a								Functional category ^b ; functional role
		Msed	Mcup	MK1	Sisl	Saci	SSO	ST	Ahos	
<i>accAB</i>	Msed_0147- Msed_0148	+	+	+	+	+	+	+	+	A: acetyl-CoA/propionyl-CoA carboxylase
<i>accC</i>	Msed_1375	+	+	+	+	+	+	+	+	A: acetyl-CoA/propionyl-CoA carboxylase
<i>mcr</i>	SSO2178	—	—	—	+	+	+	—	—	A: malonyl-CoA/succinyl-CoA reductase
<i>msr</i>	Msed_1993	+	+	+	+	+	+	+	+	A: malonate semialdehyde reductase
<i>hpcS</i>	Msed_1456	+	+	+	+	+	+	+	—	A: 3-hydroxypropionyl-CoA synthetase
<i>acr</i>	Msed_1426	+	+	—	+	+	+	+	+	A: acryloyl-CoA reductase
<i>mce</i>	Msed_0639	+	+	+	+	+	+	+	+	A: methylmalonyl-CoA epimerase
<i>mcmA</i>	Msed_0638	+	+	+	+	+	+	+	+	A: methylmalonyl-CoA mutase
<i>mcmB</i>	Msed_2055	+	+	+	+	+	+	+	+	A: methylmalonyl-CoA mutase
<i>ssr</i>	Msed_1424	+	+	+	+	+	+	+	+	A: succinate semialdehyde reductase
<i>hbcS</i>	Msed_1422	+	+	+	+	+	+	+	+	A: 4-hydroxybutyryl-CoA synthetase
<i>hbcD</i>	Msed_1321	+	+	+	+	+	+	+	+	A: 4-hydroxybutyryl-CoA dehydratase
<i>ack</i>	Msed_0656	+	+	+	+	+	+	+	+	A: acetoacetyl-CoA-ketothiolase
<i>hbd</i>	Msed_1423	+	+	+	—	0	—	—	+	A: 3-hydroxybutyryl-CoA dehydrogenase
<i>doxDA</i>	Msed_0363- Msed_0364	+	+	+	0	0	+	—	+	E: thiosulfate:quinone oxidoreductase
<i>fnr</i>	Msed_0743	+	+	+	+	+	+	+	+	E: putative ferredoxin-NADP reductase
<i>sqrI</i>	ST2485	—	+	+	—	0	—	+	+	E: putative sulfide:quinone oxidoreductase
<i>pgm</i>	Msed_0773	+	+	—	+	—	—	+	+	C: phosphoglycerate mutase
<i>porGDAB</i>	Saci_2250- Saci_2253	—	—	—	+	+	+	—	+	C: pyruvate synthase
<i>ladH</i>	SSO1629	—	—	—	+	—	+	—	+	C: lactaldehyde dehydrogenase
<i>acsA</i>	Msed_0267	+	+	+	+	—	+	+	—	C: acetyl-CoA synthetase
<i>acdB</i>	Msed_0272	+	+	—	+	—	+	—	0	F: putative acyl-CoA dehydrogenase
<i>tlsAB</i>	Msed_0270- Msed_0271	+	+	+	+	—	+	—	0	F: putative acetyl-CoA acetyltransferases
<i>amt</i>	Saci_0883	—	+	+	—	+	—	+	+	N: ammonium transporter
<i>trxA</i>	Msed_0776	+	+	+	+	—	+	+	+	Thioredoxin
<i>dhl1</i>	Msed_0527	+	+	+	—	—	—	+	—	Putative dehydrogenase
<i>dhl2</i>	Msed_0775	+	+	+	+	—	+	+	+	Putative dehydrogenase
<i>dhl3</i>	Msed_0777	+	+	+	+	—	+	+	+	Putative dehydrogenase
<i>naaB</i>	Msed_1994	+	+	+	+	+	+	+	+	Putative oxygenase
<i>hhcR</i>	Msed_2175	+	+	+	—	—	—	—	+	Transcription regulator, TrmB family
<i>hyp1-2</i>	Msed_0268- Msed_0269	+	+	+	+	—	+	—	0	Hypothetical proteins
<i>hyp3</i>	Msed_0526	+	+	+	0	0	0	—	0	Hypothetical protein
<i>hyp4</i>	Msed_2056	+	+	+	+	+	+	+	+	Putative GTPase
<i>hyp5</i>	Msed_2067	+	+	—	—	—	—	+	—	Hypothetical protein

^a The table shows the distribution of gene orthologs and candidate HhcR sites upstream of the genes across eight genomes: +, the gene ortholog is present and has an HhcR site; —, the gene is present but is not preceded by an HhcR site; 0, no gene ortholog is present in the genome. Genome abbreviations: Msed, *Metallosphaera sedula*; Mcup, *Metallosphaera cuprina*; MK1, *Metallosphaera yellowstonensis*; Sisl, *Sulfolobus islandicus*; Saci, *Sulfolobus acidocaldarius*; SSO, *Sulfolobus solfataricus*; ST, *Sulfolobus tokodaii*; Ahos, *Acidianus hospitalis*. A complete version of the table is provided as Table S2 in the supplemental material.

^b Functional categories of genes: A, autotrophic carbon dioxide fixation; E, energy metabolism; C, carbon metabolism; F, fatty acid metabolism; N, nitrogen and amino acid metabolism.

central carbon metabolism—putative acetyl-CoA acetyltransferase (*tlsAB*) and putative acyl-CoA dehydrogenase (*acdB*).

In *Metallosphaera* spp. and *A. hospitalis*, the HhcR regulons include a hypothetical TF from the COG3355 family (e.g., Msed_2175 in *M. sedula*), tentatively named HhcR. Orthologs of HhcR were found in all genomes from the order *Sulfolobales* (with >52% amino acid sequence identity) but not in other lineages of *Archaea*. HhcRs are small proteins of 110 to 120 amino acids that contain a winged helix-turn-helix DNA-binding domain from the TrmB family (Pfam ID, PF01978) but lack an effector-binding domain. Other TrmB family regulators in *Archaea*, including MreA in *Methanosarcina acetivorans*, that also lack an effector-binding domain (29) and TrmBs in the *Thermococcales* that have the effector-binding domain (30) have been characterized as global regulators of the carbon- and energy-converting pathways. We propose that HhcR functions in the global regulation of the HHC and other carbon and energy metabolism genes in the *Sulfolobales*.

Despite early reports on isolation of the *Sulfolobales* species describing most of them as autotrophs, subsequent phenotypic studies did not confirm autotrophic growth for *S. acidocaldarius* and *S. tokodaii* (31), although all the predicted HHC enzymes are present in their genomes (see Table S2 in the supplemental material). Since HhcR proteins and their DNA-binding motifs and reconstructed regulons are conserved among all *Sulfolobales*, it is unlikely that the previously observed phenotypic differences between *Sulfolobus* species are related to HhcR-dependent regulation.

Experimental validation of HhcR regulons in *Metallosphaera* spp. (i) HhcR binds the predicted HHC box motifs *in vitro*. In order to validate the computationally predicted DNA-binding motif of HhcR, we heterologously expressed and purified the HhcR protein from *M. yellowstonensis* MK-1 (8) (Uniprot ID, H2C8P4). The recombinant HhcR protein exists as a dimer in solution, supporting the hypothesis that the HhcR dimer binds to the HHC box DNA motif with a tandem-repeat structure. The

FIG 3 Experimental validation of the HhcR regulon in *M. yellowstonensis*. (A) Fluorescence polarization binding assay of HhcR with target DNA operators. Increasing protein concentrations were mixed with 28- or 37-bp fluorescence-labeled DNA fragments of gene regions containing HHC box sites. As a negative control (N.C.), a DNA fragment of the cobalt transporter gene *hoxN* from the vitamin B₁₂ regulon was used. The apparent K_d values for HhcR protein interacting with the tested DNA fragments are indicated. The error bars indicate standard deviations. (B) EMSA with an *hhcR* DNA fragment in the absence and in the presence of increasing HhcR protein concentrations. As a negative control, a DNA fragment of the thiamine biosynthesis gene *thi4* was used.

interaction of the purified HhcR proteins with their cognate DNA-binding sites was assessed using the fluorescence polarization DNA-binding assay (Fig. 3A). The results show that HhcR specifically binds to the synthetic DNA fragments containing HHC boxes at four predicted target genes in *M. yellowstonensis* (*accAB*, *trxA*, *ssr*, and *hhcR*). All tested DNA fragments demonstrated the HhcR concentration-dependent increase of fluorescence polarization, confirming specific interactions between the regulator and DNA fragments. The apparent K_d values for HhcR protein interacting with the tested DNA fragments were in the range of 135 to 380 nM (Fig. 3A). Additional EMSAs confirmed the specific binding between the recombinant HhcR protein and its cognate operator (Fig. 3B). The HhcR concentration-dependent shift of DNA bands was observed with the *hhcR* gene DNA fragment containing the predicted HHC box site, whereas the negative-control DNA fragment was not shifted. We also tested the influence of potential effectors (simple carbon sources and HHC cycle metabolites) on protein-DNA interaction; however, the addition of acetate, succinate, or acetyl-CoA had no effect on complex formation (data not shown).

(ii) Expression of HhcR-dependent genes is upregulated under autotrophic conditions. The reconstructed HhcR regulons in *M. sedula* and *M. cupina* were compared with available transcriptomics and proteomics data (see Tables S1 and S2 in the supplemental material). As result, 36 genes from the HhcR regulon, including 15 genes from the HHC pathway in *M. sedula*, were previously shown to have significantly increased expression in cells grown autotrophically compared with heterotrophically grown cells (7, 14). Thus, the HhcR regulon in *M. sedula* is consistent with previous transcriptomics analysis except for the absence of an HHC box site upstream of the malonyl-CoA/succinyl-CoA reductase gene *mcr*, which exhibited strong expression change (14) that could be explained by the effects of other regulators. In *M. cupina*, 12 proteins from the HHC pathway and 3 other proteins from the HhcR regulon have significantly increased

abundance in cells grown autotrophically versus heterotrophically (32).

Identification of the DhcR regulon in *Thermoproteales*. Regulon reconstruction for the DHC autotrophic pathway in *Thermoproteales* was started with a DNA motif search for genes in *T. neutrophilus* that were previously determined by comparative proteomic analysis and enzyme activity measurement to be autotrophically induced (15). Thus, we collected upstream regions of the *ppcA*, *ssr*, *scr-4hbl*, *4hbd-frdAB*, *ppsA*, *sucCD*, *mdh*, *hpcD*, and *fumAB* genes in *T. neutrophilus* and their orthologs in other *Thermoproteales* genomes (see Table S3 in the supplemental material). Using the MEME program, we found a conserved palindromic 21-bp motif with the consensus sequence ATTTTAnATA ATTATAAAAT (where n is any nucleotide) that was termed the DHC box (Fig. 2B). The DHC box motif was located upstream of nearly half of the selected *T. neutrophilus* operons, including *ppcA*, *ssr*, *scr-4hbl*, *4hbd-frdAB*, and *ppsA*, and was conserved in most of the 10 studied *Thermoproteales* genomes. A DHC box recognition profile was constructed and used to scan the genomes, resulting in identification of additional candidate sites upstream of 20 to 30 genes per genome. False-positive site predictions were eliminated by using the consistency check approach (see Materials and Methods). In addition, selected DHC boxes upstream of the DHC pathway genes were validated by phylogenetic footprinting (see Fig. S1B in the supplemental material). As result, the reconstructed DHC box regulons (termed DhcR) included 10 to 20 putative target operons that contained up to 40 genes per genome (see Table S4 in the supplemental material). The distribution of predicted DhcR regulon members across the *Thermoproteales* genomes is shown in Table 2.

The reconstructed DhcR regulons contain genes encoding enzymes for 8 of 14 reactions from the DHC pathway (Fig. 1). Genes encoding three different pyruvate synthases (*korAB*, *porGDAB*, and *porBAG2*), as well as phosphoenolpyruvate carboxylase (*ppcA*), succinic semialdehyde reductase (*ssr*), 4-hydroxybutyryl-

TABLE 2 Distribution of predicted DhcR regulon members across the *Thermoproteales* genomes

Gene name	Example locus tag	Presence ^a										Functional category ^b : functional role
		Pars	Pisl	Pcal	PAE	Tneu	Pogu	P186	WP30	TTX	TUZN	
<i>korAB</i>	TTX_1758–TTX_1757	+	–	–	–	–	+	+	+	+	+	A: pyruvate synthase I
<i>porGDAB</i>	Tneu_0176–Tneu_0179	+	+	–	+	+	+	–	+	+	–	A: pyruvate synthase II
<i>porGAB2</i>	Tneu_1797–Tneu_1795	+	+	+	0	+	0	+	+	+	+	A: pyruvate synthase III
<i>ppcA</i>	Tneu_0418	+	+	–	–	+	+	–	+	+	+	A: phosphoenolpyruvate carboxylase
<i>ssr</i>	Tneu_0419	–	+	–	–	+	+	+	+	+	+	A: succinic semialdehyde reductase
<i>4hbl</i>	Tneu_0420	+	+	+	+	+	+	+	+	+	+	A: 4-hydroxybutyryl-CoA synthetase
<i>scr</i>	Tneu_0421	–	+	+	–	+	+	+	+	+	+	A: succinyl-CoA reductase
<i>4hbd</i>	Tneu_0422	0	+	+	+	+	+	+	+	+	+	A: 4-hydroxybutyryl-CoA dehydratase
<i>frdAB</i>	Tneu_0423–Tneu_0424	–	+	+	–	+	–	+	+	+	+	A: fumarate reductase
<i>ppsA</i>	Tneu_1204	+	+	+	+	+	+	+	+	+	–	A: pyruvate-water dikinase
<i>nuo operon</i>	Tneu_1802–Tneu_1789	+	+	+	0	+	+	+	+	+	+	E: NADH-quinone oxidoreductase
<i>acsA</i>	Tneu_1843	+	+	+++	++	+	+	++++	++	+++	++	C: acetyl-CoA synthetase
<i>gltA</i>	TTX_0497	+	0	+	++	0	+	+	+	+	+	C: citrate synthase
<i>mae</i>	TTX_1514	+	+	+	+	0	+	+	+	+	–	C: malate dehydrogenase (malic enzyme)
<i>acaB</i>	TTX_1111	0	0	0	+	0	+	–	0	+	+	F: putative acetyl-CoA acetyltransferase
<i>metB</i>	Tneu_0243	+	+	–	+	+	+	–	–	+	–	N: cystathionine gamma-synthase
<i>gdhA</i>	TTX_1095	+	+	–	–	–	+	–	–	+	+	N: glutamate dehydrogenase
<i>livKMHGH</i>	Tneu_0549–Tneu_0545	–	+	–	+	+	+	–	–	–	–	N: branched-chain amino acid transporter
<i>atf</i>	Pcal_0578	+	+	–	+	0	+	+	+	–	–	Putative N-acetyltransferase
<i>cbs1-2</i>	Tneu_0416–Tneu_0417	0	+	–	–	+	+	–	–	+	+	CBS domain proteins
<i>cbs3</i>	TTX_0532	+	–	–	+	–	+	–	–	+	–	CBS domain protein
<i>cbs4</i>	TTX_1103	–	0	+	+	–	+	–	–	+	+	CBS domain protein
<i>hyp1</i>	Pogu_2479	+	+	–	+	–	+	+	–	–	–	Hypothetical protein
<i>hyp2</i>	Tneu_1789	+	+	+	0	+	+	0	+	0	0	Hypothetical protein
<i>hyp3</i>	TTX_0945	+	–	+	–	–	+	+	–	+	–	Hypothetical protein
<i>hyp4</i>	TTX_1112	0	0	0	+	0	+	–	0	+	+	Predicted ucleic acid-binding protein
<i>hyp5</i>	Tneu_1844	+	+	+	+	+	0	+	+	0	0	MFS transporter with CBS domain
<i>hyp6</i>	Tneu_1801	–	+	0	0	+	–	+	–	–	+	Hypothetical protein

^a +, the gene ortholog is present and has an HhcR site; –, the gene is present but is not preceded by an HhcR site; 0, no gene ortholog is present in the genome; ++, +++, and +++++, two, three, and four homologs are preceded by HhcR sites. Genome abbreviations: Pars, *Pyrobaculum arsenaticum*; Pisl, *Pyrobaculum islandicum*; Pcal, *Pyrobaculum calidifontis*; PAE, *Pyrobaculum aerophilum*; Tneu, *Thermoproteus neutrophilus*; Pogu, *Pyrobaculum oguniense*; P186, *Pyrobaculum* sp. 1860; WP30, *Pyrobaculum yellowstonensis*; WP30; TTX, *Thermoproteus tenax*; TUZN, *Thermoproteus uzoniensis*. A complete version of the table is provided as Table S3 in the supplemental material.

^b Functional categories of genes: A, autotrophic carbon dioxide fixation; E, energy metabolism; C, carbon metabolism; F, fatty acid metabolism; N, nitrogen and amino acid metabolism.

CoA synthetase (*4hbl*), succinyl-CoA reductase (*scr*), 4-hydroxybutyryl-CoA dehydratase (*4hbd*), fumarate reductase (*frdAB*), and pyruvate-water dikinase (*ppsA*), have predicted DHC boxes that are conserved in more than 6 out of 10 studied *Thermoproteales* genomes. These enzymes form two blocks of DHC reactions: (i) conversion of acetyl-CoA to oxaloacetate and (ii) conversion of fumarate to crotonyl-CoA. Interestingly, *P. arsenaticum* has an incomplete DHC pathway, as the 4-hydroxybutyryl-CoA dehydratase gene *4hbd* is missing in its genome, while several other DHC pathway genes (*ssr*, *scr*, and *frdAB*) do not belong to the reconstructed DhcR regulon. However, previous studies demonstrated that *P. arsenaticum* is able to grow chemolithoautotrophically with carbon dioxide as a carbon source (33). Additional searches among the candidate members of the DhcR regulon did not help to identify a potential *4hbd* gene substitute in the *P. arsenaticum* genome. Such a candidate potentially is not coregulated by DhcR, like several other DHC enzymes, including succinyl-CoA synthetase (*sucCD*), malate dehydrogenase (*mdh*), crotonyl-CoA hydratase (*hpcD*), and fumarate hydratase (*fumAB*), which do not belong to the reconstructed DhcR regulon in any *Thermoproteales* genome (see Table S3 in the supplemental material).

Other members of the reconstructed DhcR regulons can be classified into several functional groups (Table 2). The largest and most conserved group includes enzymes involved in carbon and energy metabolism. In all studied *Pyrobaculum* genomes, we observed multiple paralogs of acetyl-CoA synthetase (*acsA*) that are potentially coregulated by DHC boxes. *AcsA* catalyzes the conversion of acetate into acetyl-CoA, an essential intermediate at the junction of anabolic and catabolic pathways. Citrate synthase (*gltA*) and malic enzyme (*mae*) belong to the DhcR regulon in 8 out of 10 genomes. Citrate synthase catalyzes conversion of acetyl-CoA and oxaloacetate to citrate, which can be an interconnection between DHC and 2-oxoglutarate (glutamate) biosynthesis. Malic enzyme catalyzes NAD-dependent decarboxylation of malate into pyruvate. We found potential DhcR-dependent regulation of NADH-quinone oxidoreductase subunit genes (*nuo*) in all *Thermoproteales* species except *P. aerophilum*, which lacks *nuo* gene orthologs. Interestingly, the *nuo* genes were often found within the same operon with the *porGAB2* genes, encoding one of three pyruvate synthase paralogs. Coupling of the *nuo* and *porGAD2* genes suggests that the NADH-quinone oxidoreductase might be involved in ferredoxin regeneration, as the reduced ferredoxin is

involved in the reaction of carbon dioxide and acetyl-CoA condensation by the pyruvate synthase (Fig. 1).

Two genes encoding amino acid biosynthesis enzymes, namely, cystathionine gamma-synthase (*metB*) and glutamate dehydrogenase (*gdhA*), are potentially coregulated by DHC boxes in five or six genomes. These reactions are directly linked with the DHC autotrophic pathway, and thus, DHC intermediates could be used as carbon skeletons for amino acid biosynthesis. A branched-chain amino acid transporter (*ilvBHMGF*) belongs to the reconstructed DhcR regulon in four *Pyrobaculum* genomes. Finally, the predicted DhcR regulons include multiple hypothetical genes, including at least four orthologous groups of genes encoding CBS domain proteins (Table 2). The CBS domain (PF00571) is a protein domain found in a wide variety of proteins, where it plays a regulatory role in the activity of associated enzymatic domains in response to various ligands (34). This could suggest a potential role of DhcR-regulated CBS domain proteins as regulators associated with autotrophic carbon dioxide fixation cycle enzyme activities.

Previously, it was shown that a hypothetical 18-kDa protein encoded by *Tneu_0751* in *T. neutrophilus* specifically binds to the common intergenic region of the *scr-4hbl* and *4hbd-frdAB* operons (15). This regulatory region contains a high-scored DHC box that is conserved in all but one *Thermoproteales* genome (Table 2). *P. arsenaticum* lacks an ortholog of the *4hbd* gene, whereas the *scr* gene is not preceded by a DHC box in this genome. As previously observed (15), orthologs of *Tneu_0751* were identified in all studied *Thermoproteales* genomes but not in any other lineages of *Crenarchaeota* (see Table S3 in the supplemental material). We tentatively assigned to *Tneu_0751* and its orthologs the role of a cognate transcriptional regulator for the reconstructed DhcR regulon in the *Thermoproteales*. No potential DHC boxes can be identified in the upstream regions of the *dhcR* regulator genes in the *Thermoproteales*. This observation is in agreement with the previously observed constitutive expression of *Tneu_0751*, suggesting that the DhcR regulator potentially responds to some yet-unknown inductor or effector molecule.

Conclusions. Autotrophic carbon dioxide fixation pathways differ significantly between various lineages of *Archaea*. Species from the phylum *Thermoprotei* (*Crenarchaeota*) possess one of the two autotrophic pathways, HHC or DHC (Fig. 4). Before this work, little was known about the regulation of autotrophic carbon dioxide fixation in *Archaea*. By using the comparative genomics approach, we identified novel DNA-regulatory motifs associated with the autotrophic pathway genes in archaeal genomes from the *Sulfolobales* and *Thermoproteales* lineages. The HHC box motif, which is a tandem repeat of two 9-bp consensus sequences, was found in the *Sulfolobales* genomes, whereas the DHC box motif, a 21-bp AT-rich palindrome, was found in the *Thermoproteales* genomes but not in other *Crenarchaeota* (Fig. 2). Reconstruction of these global autotrophic regulons in the available genomes from two lineages revealed multiple additional genes involved in carbon and energy metabolism that are potentially coregulated with autotrophic pathway genes. The bioinformatically predicted candidate regulons were assessed using the high-throughput expression data available for several model archaeal species. The predicted function of the candidate HhcR regulator, binding the HHC box motifs in *M. yellowstonensis*, was experimentally validated by *in vitro* DNA-binding assays. The DhcR regulator was tentatively attributed to the DHC box regulon in the *Thermoproteales* based

The Thermoprotei phylum (*Crenarchaeota*)

FIG 4 Distribution of autotrophic pathways and regulons in *Crenarchaeota*. The phylogenetic species tree was constructed using the concatenated alignment of 78 universal bacterial proteins in the MicrobesOnline database (<http://www.microbesonline.org/cgi-bin/speciesTree.cgi>). *P. arsenaticum* lacks the key enzyme of the pathway, *4hbd*; however, despite the fact that its DHC pathway is incomplete, *P. arsenaticum* was reported to be able to grow chemolithoautotrophically with carbon dioxide as a carbon source.

on the previous DNA affinity purification studies in *T. neutrophilus*. The exact roles of HhcR and DhcR regulators in their predicted transcriptional regulons are yet to be elucidated using the genetic and/or transcriptomics approaches. In summary, this study provides novel insights into transcriptional regulation of autotrophic pathways in two lineages of *Crenarchaeota*.

ACKNOWLEDGMENTS

This research was supported by the Russian Science Foundation (grant number 14-14-00289). Additional funding was provided by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

We thank Margie F. Romine for providing the tables of orthologous genes for the *Sulfolobales* and *Thermoproteales* genomes and Andrei L. Osterman for useful discussions on experimental assessment of the HhcR regulator.

REFERENCES

- Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hugler M, Alber BE, Fuchs G. 2010. Autotrophic carbon fixation in archaea. *Nat Rev Microbiol* 8:447–460. <http://dx.doi.org/10.1038/nrmicro2365>.
- Hugler M, Huber H, Stetter KO, Fuchs G. 2003. Autotrophic CO₂ fixation pathways in archaea (Crenarchaeota). *Arch Microbiol* 179: 160–173. <http://dx.doi.org/10.1007/s00203-002-0512-5>.
- Berg IA, Kockelkorn D, Buckel W, Fuchs G. 2007. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* 318:1782–1786. <http://dx.doi.org/10.1126/science.1149976>.
- Ramos-Vera WH, Berg IA, Fuchs G. 2009. Autotrophic carbon dioxide assimilation in Thermoproteales revisited. *J Bacteriol* 191:4286–4297. <http://dx.doi.org/10.1128/JB.00145-09>.
- Han Y, Hawkins AS, Adams MW, Kelly RM. 2012. Epimerase (Msd_0639) and mutase (Msd_0638 and Msd_2055) convert (S)-methylmalonyl-coenzyme A (CoA) to succinyl-CoA in the Metallosphaera sedula 3-hydroxypropionate/4-hydroxybutyrate cycle. *Appl Environ Microbiol* 78:6194–6202. <http://dx.doi.org/10.1128/AEM.01312-12>.
- Hawkins AB, Adams MW, Kelly RM. 2014. Conversion of 4-hydroxybutyrate to acetyl coenzyme A and its anapleurosis in the Metallosphaera sedula 3-hydroxypropionate/4-hydroxybutyrate carbon fixation pathway. *Appl Environ Microbiol* 80:2536–2545. <http://dx.doi.org/10.1128/AEM.04146-13>.
- Hawkins AS, Han Y, Bennett RK, Adams MW, Kelly RM. 2013. Role of 4-hydroxybutyrate-CoA synthetase in the CO₂ fixation cycle in thermoacidophilic archaea. *J Biol Chem* 288:4012–4022. <http://dx.doi.org/10.1074/jbc.M112.413195>.
- Jennings RM, Whitmore LM, Moran JJ, Kreuzer HW, Inskeep WP. 2014. Carbon dioxide fixation by Metallosphaera yellowstonensis and acidothermophilic iron-oxidizing microbial communities from Yellowstone National Park. *Appl Environ Microbiol* 80:2665–2671. <http://dx.doi.org/10.1128/AEM.03416-13>.
- Ramos-Vera WH, Weiss M, Strittmatter E, Kockelkorn D, Fuchs G. 2011. Identification of missing genes and enzymes for autotrophic carbon fixation in crenarchaeota. *J Bacteriol* 193:1201–1211. <http://dx.doi.org/10.1128/JB.01156-10>.
- Berg IA, Ramos-Vera WH, Petri A, Huber H, Fuchs G. 2010. Study of the distribution of autotrophic CO₂ fixation cycles in Crenarchaeota. *Microbiology* 156:256–269. <http://dx.doi.org/10.1099/mic.0.034298-0>.
- Estelmann S, Hugler M, Eisenreich W, Werner K, Berg IA, Ramos-Vera WH, Say RF, Kockelkorn D, Gad'on N, Fuchs G. 2011. Labeling and enzyme studies of the central carbon metabolism in Metallosphaera sedula. *J Bacteriol* 193:1191–1200. <http://dx.doi.org/10.1128/JB.01155-10>.
- Ishii M, Chuakrut S, Arai H, Igarashi Y. 2004. Occurrence, biochemistry and possible biotechnological application of the 3-hydroxypropionate cycle. *Appl Microbiol Biotechnol* 64:605–610. <http://dx.doi.org/10.1007/s00253-003-1540-z>.
- Hugler M, Krieger RS, Jahn M, Fuchs G. 2003. Characterization of acetyl-CoA/propionyl-CoA carboxylase in Metallosphaera sedula. Carboxylating enzyme in the 3-hydroxypropionate cycle for autotrophic carbon fixation. *Eur J Biochem* 270:736–744. <http://dx.doi.org/10.1046/j.1432-1033.2003.03434.x>.
- Auernik KS, Kelly RM. 2010. Physiological versatility of the extremely thermoacidophilic archaeon Metallosphaera sedula supported by transcriptomic analysis of heterotrophic, autotrophic, and mixotrophic growth. *Appl Environ Microbiol* 76:931–935. <http://dx.doi.org/10.1128/AEM.01336-09>.
- Ramos-Vera WH, Labonte V, Weiss M, Pauly J, Fuchs G. 2010. Regulation of autotrophic CO₂ fixation in the archaeon Thermoproteus neutrophilus. *J Bacteriol* 192:5329–5340. <http://dx.doi.org/10.1128/JB.00729-10>.
- Leyn SA, Rodionov DA. 2015. Comparative genomics of DtxR family regulons for metal homeostasis in Archaea. *J Bacteriol* 197:451–458. <http://dx.doi.org/10.1128/JB.02386-14>.
- Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. 2012. GenBank. *Nucleic Acids Res* 40:D48–D53. <http://dx.doi.org/10.1093/nar/gkr1202>.
- Macur RE, Jay ZJ, Taylor WP, Kozubal MA, Kocar BD, Inskeep WP. 2013. Microbial community structure and sulfur biogeochemistry in mildly-acidic sulfidic geothermal springs in Yellowstone National Park. *Geobiology* 11:86–99. <http://dx.doi.org/10.1111/gbi.12015>.
- Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC. 2014. IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res* 42:D560–D567. <http://dx.doi.org/10.1093/nar/gkt963>.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <http://dx.doi.org/10.1093/nar/25.17.3389>.
- Mironov A, Vinokurova N, Gelfand M. 2000. Software for analysis of bacterial genomes. *Mol Biol* 34:222–231. <http://dx.doi.org/10.1007/BF02759643>.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <http://dx.doi.org/10.1093/nar/gkh340>.
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37:W202–W208. <http://dx.doi.org/10.1093/nar/gkp335>.
- Gelfand MS, Koonin EV, Mironov AA. 2000. Prediction of transcription regulatory sites in Archaea by a comparative genomic approach. *Nucleic Acids Res* 28:695–705. <http://dx.doi.org/10.1093/nar/28.3.695>.
- Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. *Nucleic Acids Res* 40:D290–D301. <http://dx.doi.org/10.1093/nar/gkr1065>.
- Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS, Dubchak IL, Alm EJ, Arkin AP. 2010. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res* 38:D396–D400. <http://dx.doi.org/10.1093/nar/gkp919>.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190. <http://dx.doi.org/10.1101/gr.849004>.
- Rodionov DA. 2007. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem Rev* 107:3467–3497. <http://dx.doi.org/10.1021/cr068309+>.
- Reichlen MJ, Vepachedu VR, Murakami KS, Ferry JG. 2012. MreA functions in the global regulation of methanogenic pathways in Methanosarcina acetivorans. *mBio* 3:e00189-00112. <http://dx.doi.org/10.1128/mBio.00189-12>.
- Gindner A, Hausner W, Thomm M. 2014. The TrmB family: a versatile group of transcriptional regulators in Archaea. *Extremophiles* 18:925–936. <http://dx.doi.org/10.1007/s00792-014-0677-2>.
- Grogan DW. 1989. Phenotypic characterization of the archaeobacterial genus Sulfolobus: comparison of five wild-type strains. *J Bacteriol* 171: 6710–6719.
- Jiang CY, Liu LJ, Guo X, You XY, Liu SJ, Poetsch A. 2014. Resolution of carbon metabolism and sulfur-oxidation pathways of Metallosphaera cuprina Ar-4 via comparative proteomics. *J Proteomics* 109C:276–289. <http://dx.doi.org/10.1016/j.jprot.2014.07.004>.
- Huber R, Sacher M, Vollmann A, Huber H, Rose D. 2000. Respiration of arsenate and selenate by hyperthermophilic archaea. *Syst Appl Microbiol* 23:305–314. [http://dx.doi.org/10.1016/S0723-2020\(00\)80058-2](http://dx.doi.org/10.1016/S0723-2020(00)80058-2).
- Ereno-Orbea J, Oyenarte I, Martinez-Cruz LA. 2013. CBS domains: ligand binding sites and conformational variability. *Arch Biochem Biophys* 540:70–81. <http://dx.doi.org/10.1016/j.abb.2013.10.008>.