JOURNAL OF BACTERIOLOGY, Nov. 2004, p. 7754–7762 Vol. 186, No. 22
0021-9193/04/$08.00+0 DOI: 10.1128/JB.186.22.7754–7762.2004


# Identification and Functional Verification of Archaeal-Type Phosphoenolpyruvate Carboxylase, a Missing Link in Archaeal Central Carbohydrate Metabolism

Thijs J. G. Ettema,[1]* Kira S. Makarova,[2] Gera L. Jellema,[1] Hinco J. Gierman,[1] Eugene V. Koonin,[2] Martijn A. Huynen,[3] Willem M. de Vos,[1] and John van der Oost[1]

*Laboratory of Microbiology, Agrotechnology and Food Sciences, Wageningen University, Wageningen,[1] and Nijmegen Center for Molecular Life Sciences, Center for Molecular and Biomolecular Informatics, Nijmegen,[3] The Netherlands, and National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland[2]*

Received 10 May 2004/Accepted 16 August 2004

**Despite the fact that phosphoenolpyruvate carboxylase (PEPC) activity has been measured and in some cases even purified from some *Archaea*, the gene responsible for this activity has not been elucidated. Using sensitive sequence comparison methods, we detected a highly conserved, uncharacterized archaeal gene family that is distantly related to the catalytic core of the canonical PEPC. To verify the predicted function of this archaeal gene family, we cloned a representative from the hyperthermophilic acidophile *Sulfolobus solfataricus* and functionally produced the corresponding enzyme as a fusion with the *Escherichia coli* maltose-binding protein. The purified fusion protein indeed displayed highly thermostable PEPC activity. The structural and biochemical properties of the characterized archaeal-type PEPC (atPEPC) from *S. solfataricus* are in good agreement with previously reported biochemical analyses of other archaeal PEPC enzymes. The newly identified atPEPC, with its distinct properties, constitutes yet another example of the versatility of the enzymes of the central carbon metabolic pathways in the archaeal domain.**

Phosphoenolpyruvate carboxylase (PEPC; EC 4.1.1.31) catalyzes the irreversible β-carboxylation of phosphoenolpyruvate (PEP) to form oxaloacetate (OAA) and inorganic phosphate with $HCO_3^-$ as a cosubstrate and divalent metal ions as cofactors. PEPC is a cytosolic enzyme that is widely distributed among a great variety of *Bacteria*, as well as higher and lower plants, and performs diverse biological functions (5). In particular, PEPC plays an important role in heterotrophic bacteria and plants via anaplerotic replenishing of C4-dicarboxylic acid into the tricarboxylic acid (TCA) cycle. The synthesis of OAA is a key step in the formation of four-carbon compounds, such as malate, fumarate, and succinate. As such, PEPC performs a central role by maintaining the continuity of the TCA cycle carbon fluxes, connecting glycolysis to the TCA cycle.

PEPC orthologs are present in most bacterial lineages (COG2352, http://www.ncbi.nlm.nih.gov/COG/new/release/cow.cgi?cog = COG2352; Cluster of Orthologous Genes [COG] database) (30) and are widespread in plants. In contrast, no PEPC homologs have been detected in *Archaea* thus far. However, PEPC activities have been measured (34), and the corresponding enzymes were purified and biochemically characterized in two archaeal hyperthermophiles: *Methanothermus sociabilis* (26) and *Sulfolobus acidocaldarius* (25). In the latter studies, it has been shown that the archaeal enzyme resembles the bacterial PEPC in quaternary structure, being a tetramer that requires $Mg^{2+}$ for its activity. However, unlike the bacterial enzyme, the archaeal PEPC is considerably smaller in size and lacks some typical regulatory properties (25, 26). Interestingly, a recent comparative genomic study of central carbon metabolic pathways of autotrophic methanogens failed to identify a gene encoding an apparent candidate to connect pyruvate metabolism to the partial reductive TCA cycle in the genome of *Methanopyrus kandleri* (29). Obviously, this gene remained to be detected.

Here we describe the computational prediction of a novel, divergent archaeal-type PEPC (atPEPC) family, which is encoded in most of the sequenced archaeal genomes, including *Methanopyrus kandleri*, and the functional characterization of atPEPC from *Sulfolobus solfataricus* P2. The properties of the recombinantly produced atPEPC closely resembled the characteristics of the PEPC enzyme that was previously purified from the closely related species *S. acidocaldarius*.

## MATERIALS AND METHODS

**Database searches.** In order to detect sequences that are distantly related to the bacterial and eukaryote PEPC sequences in the archaeal domain, the Non-Redundant database (NRDB) of protein sequences (National Center for Biotechnology Information, National Institutes of Health, Bethesda, Md. [http://www.ncbi.nih.gov/BLAST]) was searched by using the PSI-BLAST program, by using the BLOSUM62 matrix and a cutoff of E = 0.01, for inclusion of sequences in position-specific scoring matrices (2, 3).

**Multiple alignments and phylogenetic tree construction.** Multiple alignments of protein sequences were constructed by using the T-Coffee program (24), followed by manual adjustment for conserved motifs based on the PSI-BLAST results. Protein secondary structure was predicted by using the JPRED program (6). The same alignment was used for phylogenetic tree reconstruction as follows. Evolutionary distances for the tree reconstructions were calculated from multiple sequence alignments by using the Dayhoff PAM model as implemented in the PROTDIST program of the PHYLIP package (8). Distance trees were

* Corresponding author. Mailing address: Laboratory of Microbiology, Wageningen University, Hesselink van Suchtelenweg 4, 6703 CT Wageningen, The Netherlands. Phone: 31(0)317-483110. Fax: 31(0)317-483829. E-mail: thijs.ettema@wur.nl.

constructed by using the least-squares method (9) as implemented in the FITCH program of PHYLIP (8). Maximum-likelihood trees were constructed by using the ProtML program of the MOLPHY package, with the JTT-F model of amino acid substitutions (1, 12), to optimize the least-square trees with local rearrangements. Bootstrap analysis was performed for the maximum-likelihood tree as implemented in MOLPHY by using the RELL (for resampling of estimated log likelihoods) method (1, 15).

**Organisms and growth conditions.** *Escherichia coli* JM109 was used both for cloning and for subsequent functional expression of target genes and was cultivated in Luria-Bertani medium containing ampicillin (100 μg/ml) in a rotary shaker at 37°C.

**Cloning and functional expression of atPEPC from *S. solfataricus*.** The method described above identified an atPEPC candidate in the genome of *S. solfataricus*, open reading frame (ORF) SSO2256 (gi|15899028, SSO-PEPC), which subsequently was chosen for recombinant expression in *E. coli*. In order to create a maltose-binding protein (MBP)–atPEPC fusion protein, ORF SSO2256 was PCR amplified with the primer pair BG1586 (5′-GCGCG**GAATTC**ATGAGAA TCATACCACGCACTATGTC-3′) and BG1587 (5′-GCGCG**GTCGAC**TCATC CCAAGGATCTTCTAATTAATGC-3′), with the introduced EcoRI and SalI restriction sites indicated in boldface. The resulting PCR product was digested and cloned into the MBP fusion vector pMAL-c2T (7), which was kindly provided by P. Riggs (New England Biolabs, Boston, Mass.). The resulting plasmid, pWUR140, was used for recombinant expression in *E. coli* JM109 by standard procedures. A 1-liter culture of *E. coli* JM109 harboring the pWUR140 plasmid was grown, and expression was at an induced optical density at 595 nm of 0.5 by addition of IPTG (isopropyl-β-D-thiogalactopyranoside; final concentration, 0.4 mM). After an additional incubation at 37°C for 4 h, allowing expression of heterologous protein, cells were harvested by centrifugation (2,200 × *g* for 10 min) and resuspended in 10 ml of column buffer (50 mM Tris-HCl [pH 8.0], 200 mM NaCl, 1 mM EDTA). Cells were then lysed by sonication (Branson), during which the sample was cooled on ice-ethanol slurry. Cell debris was removed by centrifugation (10,000 × *g* for 20 min). The resulting supernatant was used for purification of the recombinant atPEPC.

**Purification of recombinant atPEPC.** Cell extract from a 1-liter expression culture was applied to 1 ml (bed volume) of amylose resin (New England Biolabs), which was preequilibrated with 5 column volumes of column buffer. Subsequently, the amylose resin was washed with 10 column volumes of column buffer to wash out unbound proteins. Finally, highly purified, active MBP-atPEPC fusion protein could be eluted from the amylose resin by adding 1 column volume of elution buffer (column buffer containing 10 mM maltose). Routinely, a 1-liter expression culture yielded ~0.5 mg of purified MBP-atPEPC.

**Protein concentration and purity.** Protein concentrations were determined with Coomassie brilliant blue G-250 as described before (4) with bovine serum albumin as a standard. The purity of the fusion protein was checked by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Protein samples for SDS-PAGE were heated for 5 min at 100°C in an equal volume of sample loading buffer (0.1 M citrate-phosphate buffer, 5% SDS, 0.9% 2-mercaptoethanol, 20% glycerol [pH 6.8]).

**Standard PEPC enzyme assay.** PEPC activity was routinely measured in a coupled enzyme assay using the thermostable malate dehydrogenase (MDH) from *Thermus flavus* (Sigma) by spectophotometrically monitoring the NADH oxidation at 340 nm at a Hitachi 2010 spectrophotometer at 80°C. All enzyme assays, unless stated otherwise, were started by adding the substrate (PEP) to a final concentration of 5 mM to the preheated reaction mix, containing purified recombinant *S. solfataricus* MBP-atPEPC fusion protein (20 μg/ml), 50 mM Tris-HCl (pH 8.0), 10 mM $Na_2CO_3$, 2 mM $MgSO_4$, 0.15 mM NADH, and 2 U of *T. flavus* MDH in a final volume of 1 ml. One unit of PEPC activity was defined as the amount of enzyme that is needed for the conversion of 1 μmol of NADH per min.

**Effects of metal ion concentration, potential allosteric effectors of atPEPC activity.** Known allosteric activators (glucose 6-phosphate and fructose 1,6-bisphosphate) and inhibitors (L-aspartate, L-malate, and acetyl coenzyme A) of bacterial/eukaryal-type PEPC (BE-PEPC) were tested on the activity of the *S. solfataricus* atPEPC by adding various amounts (0 to 5 mM) to the standard enzyme assays at 80°C. Metal ion dependence of atPEPC activity was determined by adding different concentrations (0 to 10 mM) of either $MgSO_4$ or $MnSO_4$ to the standard enzyme assay mixture lacking divalent metal ions.

## RESULTS AND DISCUSSION

**Computational identification of the archaeal homologs of PEPC.** In the course of a systematic analysis of archaeal central carbon metabolic pathways (11, 17, 18, 33), we detected significant similarity between archaeal proteins from COG1892 and the BE-PEPC family (COG2352) by using the PSI-BLAST program (2). For example, a PSI-BLAST search that was seeded with *S. solfataricus* sequence SSO2256 (gi|15899028) from COG1892, retrieved a PEPC-like ortholog from *Thermus* sp. (gi|1581933), at the second iteration with an E value of $10^{-6}$. A reverse search initiated with the PEPC of *E. coli* (gi|20664390) retrieves PF1975 (gi|18978347), the member of COG1892 from *Pyrococcus furiosus*, with an E value $4 \times 10^{-4}$ at the second iteration, further supporting the connection between these protein families. In addition to species listed in COG1892, we detected orthologous proteins encoded in the genomes of several archaeal species (*P. furiosus*, *Methanosarcina barkeri*, *Methanosarcina mazei*, *Ferroplasma acidarmanus*, *Sulfolobus tokodaii*, and *Picrophilus torridus*; Table 1) and three gram-positive bacterial species (*Clostridium perfringens*, *Oenococcus oeni*, and *Leuconostoc mesenteroides*). Most likely, these bacteria have acquired the archaeal PEPC gene via lateral gene transfer, since these bacterial sequences confidently cluster within the archaeal sequences in a maximum likelihood-based phylogenetic tree (see Fig. 1). This seems to be a case of nonorthologous gene displacement (16) in which the archaeal enzyme apparently has displaced the typical bacterial PEPC. It should be noted that proteins belonging to this orthologous group have been hypothesized to be PEPC previously (11).

**Sequence analysis of atPEPC.** The most notable difference between BE-PEPC and atPEPC is the apparent molecular mass of the respective proteins. The mass of a typical BE-PEPC ranges from 90 to 110 kDa, whereas the calculated molecular masses of atPEPC subunits are approximately half this size, ranging from 55 to 60 kDa, a finding in good agreement with the data for the archaeal PEPCs that have been purified and biochemically characterized (60 ± 5 kDa) (25, 26). Analysis of the multiple alignment of the two families shows that the difference in size between BE-PEPC and atPEPC is mainly due to large insertions in BE-PEPC in the regions surrounding motif I (Fig. 2). According to the resolved structures of the *E. coli* and maize PEPCs, these insertions form two four-helical bundles, which are located at the dimer interface and therefore are thought to play a role in the stabilization of the tetramer (13, 14, 20). As has been observed in the maize and *E. coli* PEPC structures, these inserted regions constitute a binding site of a sulfate ion, which was located in the proposed binding site of BE-PEPC allosteric activator, glucose 6-phosphate (14, 20). Indeed, three of four conserved arginine residues (R183, R184, R231, and R372) of bacterial and eukaryotic PEPCs, which are notably absent in the archaeal sequences, have been shown to be involved in allosteric regulation by glucose 6-phosphate, emphasizing the importance of this region in modulating the enzymatic activity (14). L-Aspartate is a well-known allosteric inhibitor of BE-PEPC, and the residues R587, K773, R832, and Q881 have been identified as its binding pocket (13). Of the four amino acids responsible for L-aspartate binding, only the catalytically important R587 (motif VI) is conserved in atPEPC. In the *E. coli* PEPC, L-aspartate binding renders the enzyme inactive by dislocating R587. Two other residues, K773 and R832 of PEPC from *E. coli*, are located in the region between motifs X and XI and are only conserved throughout the BE-PEPC family.

TABLE 1. Presence of OAA-forming and -consuming enzymes encoded by archaeal genomes[a]

| Domain and species | PCK A/B COG1274/1866[b] | PYC/OAD 5016/0439[c] | atPEPC 1892 | MDH[d] 0039/2055 | CTS 0372 | CTL[f] 2301 |
|---|---|---|---|---|---|---|
| *Crenarchaea* | | | | | | |
| *Aeropyrum pernix* | −/APE0033 | −/− | - | APE0672/− | APE1713 | APE0311 |
| *Pyrobaculum aerophilum* | −/− | −/− | PAE3416 | PAE2370/− | PAE3585 + PAE3586[e] PAE1689 | − |
| *Sulfolobus solfataricus* | SSO2537/− | −/− | SSO2256 | SSO2585/− | SSO2589 | SSO1254 |
| *Sulfolobus tokodaii* | ST1058/− | −/− | ST2101 | ST1811/− | ST1805, ST0587 | − |
| *Euryarchaea* | | | | | | |
| *Archaeoglubus fulgidus* | −/− | AF1252/AF0220 | AF1486 | AF0855/− | AF1340 | − |
| *Halobacterium* sp. | −/− | −/− | VNG2259C | VNG2367G/− | VNG2102G | VNG0627G |
| *Methanosarcina acetivorans* | −/− | MA0674/MA0675 | MA2690 | MA0819/− | MA0249 | − |
| *Methanosarcina barkeri* | −/− | Meth02001244/Meth02001245 | Meth02002803 | Meth02003982/− | Meth02000729 | − |
| *Methanosarcina mazei* | −/− | MM1827/MM1828 | MM3212 | MM1966/− | MM1527 | − |
| *Methanococcus jannaschii* | −/− | MJ1231/MJ1229 | - | MJ0490/MJ1425 | - | − |
| *Methanopyrus kandleri* | −/− | −/− | MK0190 | MK1069/MK0392 | - | − |
| *Methanothermobacter thermoautotrophicus* | −/− | MTH1107/MTH1917 | MTH943 | MTH188/MTH1205 | MTH1726, MTH962 | − |
| *Pyrococcus abyssi* | PAB1253/− | −/− | PAB2342 | −/PAB1791 | - | − |
| *Pyrococcus furiosus* | PF0289/− | −/− | PF1975 | −/− | PF0203 | − |
| *Pyrococcus horikoshii* | PH0312/− | −/− | PH0016 | −/PH1277 | - | − |
| *Ferroplasma acidarmanus* | Faci02001897/− | −/− | Faci02001756 | Faci02000699/− | Faci02000612, Faci02001586 | − |
| *Picrophilus torridus* | −/− | −/− | PTO0964 | PTO0994/− | PTO0889, PTO0169 | − |
| *Thermoplasma acidophilum* | TA0123/− | −/− | - | TA0952/− | TA0169, TA0819 | − |
| *Thermoplasma vulcanium* | TVN0200/− | −/− | - | TVN1097/− | TVN0239 | − |

[a] For enzyme abbreviations, see Fig. 4. The absence of a gene ortholog is depicted with a minus sign (−).
[b] COG1274: ATP-dependent PEP carboxykinase; COG1866: GTP-dependent PEP carboxykinase.
[c] The archaeal pyruvate carboxylase/OAA decarboxylase consists of a heterodimeric protein (23): COG5016 contains the enzymatic α- subunit, and COG0439 contains the biotinylated β-subunit. If both the enzymatic (COG5016) and the biotinylated subunit (COG0439) are present, the corresponding protein identifiers are indicated.
[d] COG0039: bacterial-type MDH; COG2055 archaeal-type MDH (11).
[e] PAE3585 and PAE3586 appear to encode a single, full-length citrate synthase gene that is disrupted by a genuine frameshift.
[f] COG2301 encodes the β-subunit of the ATP:citrate lyase. Homologs of the α and γ subunits are not encoded by archaeal genomes.

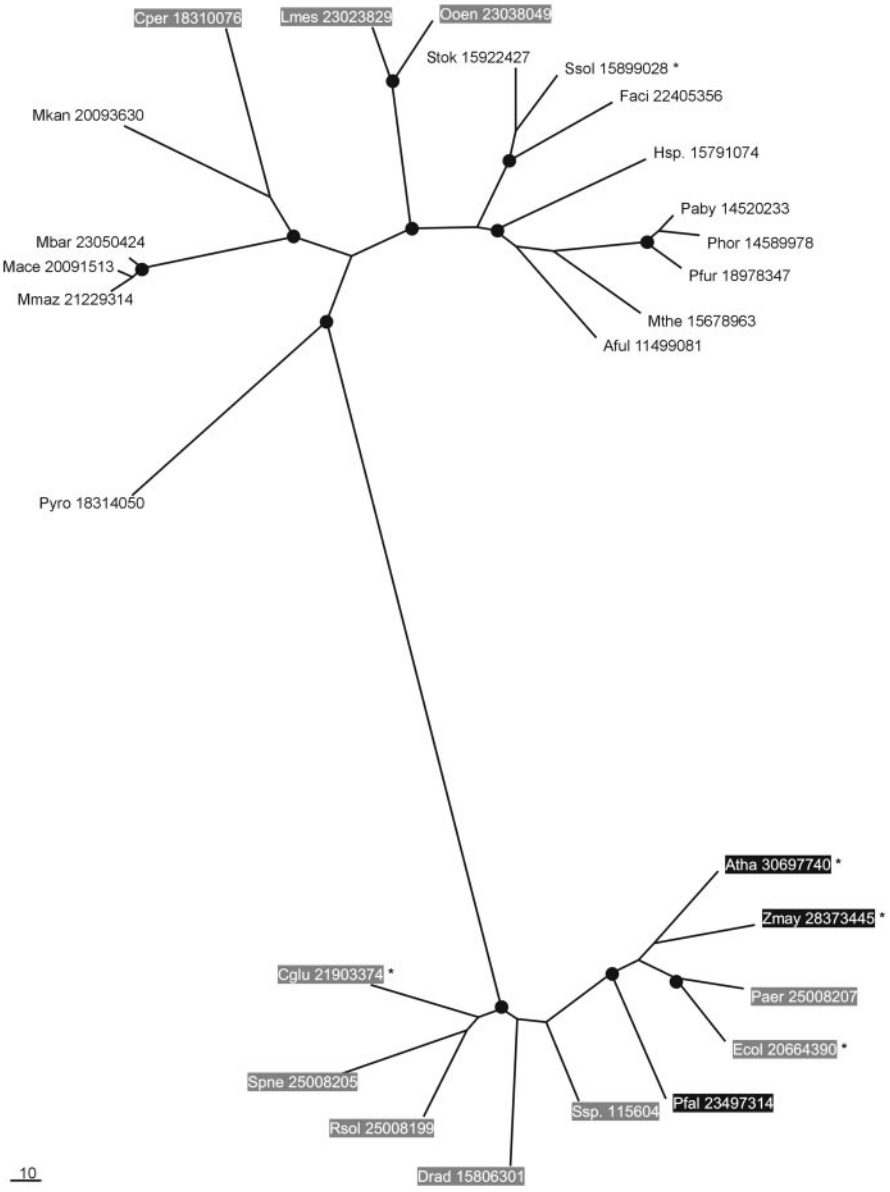


FIG. 1. Maximum-likelihood tree of atPEPC and BE-PEPC sequences. Nodes with bootstrap probability >70% are marked by circles. Archaeal proteins are in plain text, eukaryotic proteins are shaded black, and bacterial proteins are shaded gray. The sequences are denoted by their gi numbers and abbreviated species names. Species abbreviations: Pyro, *Pyrobaculum aerophilum*; Ssol, *Sulfolobus solfataricus*; Stok, *Sulfolobus tokodaii*; Hsp, *Halobacterium* sp.; Mace, *Methanosarcina acetivorans*; Mmaz, *Methanosarcina mazei*; Mbar, *Methanosarcina barkeri*; Mkan, *Methanopyrus kandleri*; Mthe, *Methanothermobacter thermautotrophicus*; Paby, *Pyrococcus abyssi*; Pfur, *Pyrococcus furiosus*; Phor, *Pyrococcus horikoshii*; Aful, *Archaeoglobus fulgidus*; Faci, *Ferroplasma acidarmanus*; Atha, *Arabidopsis thaliana*; Zmay, *Zea mays*; Pfal, *Plasmodium falciparum*; Ooen, *Oenococcus oeni*; Cper, *Clostridium perfringens*; Lmes, *Leuconostoc mesenteroides*; Paer, *Pseudomonas aeruginosa*; Ecol, *Escherichia coli*; Ssp, *Synechococcus* sp.; Drad, *Deinococcus radiodurans*; Rsol, *Ralstonia solanacearum*; Spne, *Streptococcus pneumoniae*; Cglu: *Corynebacterium glutamicum*. An asterisk marks characterized proteins.

There are no conserved residues in the archaeal family in this region (data not shown).

The rest of the protein is tightly packed around an eight-strand β-barrel and includes the active site and the PEP-binding motif (13, 19, 20). In the SCOP database (23), this domain is classified as a TIM barrel fold, superfamily PEP/pyruvate domain, which is also found in pyruvate kinase, pyruvate phosphate dikinase, and a few other enzymes that are capable of converting structurally similar substrates (http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.d.b.bd.html). Seven of the eight conserved β-strands of the BE-PEPC family could be confidently identified in atPEPC (see motifs I to VII in Fig. 2), with the sole exception of strand-2. However, the atPEPC family proteins have a predicted β-strand between motifs I and II that could potentially occupy the position of strand-2 in the atPEPC structure (Fig. 2). Most of the catalytically important amino acid residues of BE-PEPCs are conserved in atPEPCs, except for the single substitution of a conserved arginine residue that

FIG. 2. Multiple alignment of the conserved core of archaeal, bacterial and eukaryote PEPC sequences. The sequences are denoted by gene identification (gi) numbers from the GenBank database, species abbreviations (see the legend to Fig. 1), and systematic gene numbers; proteins with available structures are denoted by their PDB code. The positions of the first and last residues of the aligned region in the corresponding protein are indicated for each sequence. The numbers within the alignment represent poorly conserved inserts that are not shown. Amino acids residues that are involved in the formation of the active center are shown by asterisks; those that have an anticipated involvement in L-aspartate binding are denoted as "A" in the line between atPEPC and BE-PEPC families. Positions with identical amino acids in both families are in boldface. The coloring is based on the consensus (calculated for all sequences in the alignment) shown underneath the alignment. h, Hydrophobic residues (ACFILMVWYH); t, turn-forming residues (ASTDNVGPENRK); p, charged residues (STEDKRNQH); s, small residues (AGSVC); a, aromatic residues (FYW). The secondary structure elements correspond to those experimentally identified for 1QB4 (14). H, α-helix; E, β-strand. Sequence region of β-strand 2 for 1QB4 and predicted by JPRED program (6) for gi|18310076 (CPE1094) as a query β-strand 2 for atPEPC are shown in blue.

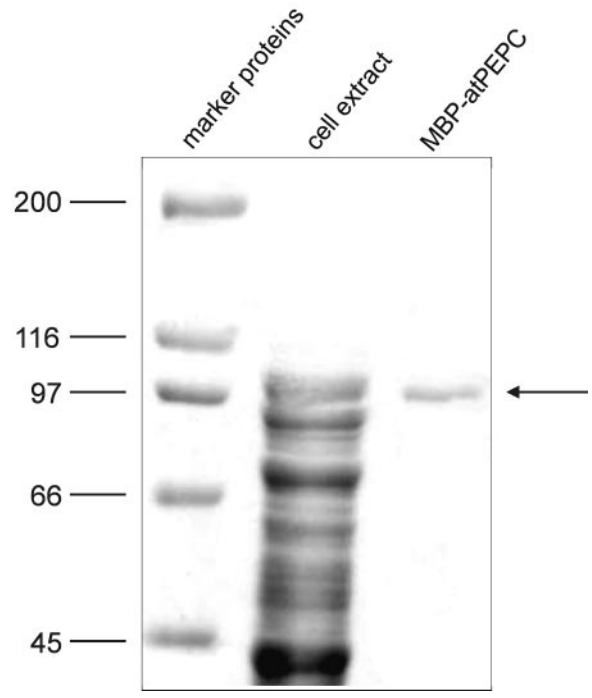


FIG. 3. SDS-PAGE of purified of MBP-atPEPC fusion protein. The SDS-PAGE was performed at a polyacrylamide concentration of 8% and with either 500 μg of cell extract of an expression culture or 10 μg of purified fusion protein, which was obtained by using affinity chromatography (see Materials and Methods). The arrow indicates a single band of ~100 kDa, closely corresponding to the calculated molecular mass of the fusion protein (101 kDa).

is replaced by a glutamic acid in motif II of the archaeal family (Fig. 2).

Taken together, the results of these computational analyses suggest that the archaeal homologs of BE-PEPCs have the same enzymatic activity but are unlikely to be subject to the same type of allosteric regulation.

**Heterologous production and purification of atPEPC.** Extensive effort to functionally produce several atPEPC candidates from different archaeal species using T7 expression systems were unsuccessful due to the formation of inclusion bodies (not shown). Subsequently, an attempt was made to enhance the solubility by producing atPEPC fused to the *E. coli* MBP. The ability of MBPs to enhance the solubility of similar "hard-case" proteins has been well studied (10). In order to create a MBP fusion protein, ORF SSO2256 from *S. solfataricus*, encoding the atPEPC candidate, was PCR amplified from genomic DNA from *S. solfataricus* P2 and cloned into the *E. coli* MBP fusion vector pMAL-c2T, resulting in pWUR140. The atPEPC candidate from *S. solfataricus* was functionally expressed as an MBP fusion protein and could be purified from cleared lysates from an expression culture by affinity chromatography using an amylose resin (New England Biolabs) (Fig. 3A). After elution from amylose resin with elution buffer, the MBP-atPEPC fusion protein eluted as a single band with an estimated molecular mass of ca. 100 kDa according to SDS-PAGE analysis. Several attempts to separate MBP from the atPEPC enzyme by proteolytic cleavage by using the thrombin site that is present in the linker region in between the

TABLE 2. Comparison of characterized at BEPCs and BE-PEPCs

| Parameter | Characteristics of various species[c] | | | | |
|---|---|---|---|---|---|
| | *Archaea* | | | *Bacteria* (*E. coli*) | *Eukarya* (*Z. mays*) |
| | *S. solfataricus* | *S. acidocaldarius* | *M. sociabilis* | | |
| Mass (kDa)/no. of subunits | ND[a] | 260/4 | 240/4 | 360/4 | 400/4 |
| Mass (kDa)/monomer | 58[b] | 60 | 60 | 90 | 100 |
| $T_{opt}$ | 85 | 90 | 85 | 35–38 | 40 |
| $K_m$ (PEP) | 0.09 | 0.2 | 1.3 | 20 | 0.1 |
| $V_{max}$ (U/mg) | 2.1 | 2.77 | 2.70 | 30 | 20.8 |
| Optimum pH | 8.0 | 8.0 | 8.5 | 7.5 | 7.5–8.0 |
| Allosteric inhibitors | L-Aspartate, L-malate | L-Aspartate, L-malate | | L-Aspartate, L-malate, citrate, succinate, fumarate | L-Aspartate, L-malate, citrate, pyruvate, oxaloacetate G6P, G1P, glycine |
| Allosteric activators | | | | Ac-CoA, FBP, fatty acids, GTP | |
| Metal(s) | $Mg^{2+}$ | $Mg^{2+}$ | $Mg^{2+}$, $Mn^{2+}$ | $Mg^{2+}$ | $Mg^{2+}$ |
| Source or reference | This study | 25 | 26 | 31 | 32 |

[a] The molecular mass and multimeric conformation of the native MBP-PEPC could not be determined accurately (ND) since the fusion protein was found to be associated in a large complex (T. J. G. Ettema, unpublished observation).
[b] Calculated molecular mass.
[c] Abbreviations: Ac-CoA, acetyl coenzyme A; FBP, fructose 1,6-bisphosphate; G6P, glucose 6-phosphate; G1P, glucose 1-phosphate.

fused proteins were unsuccessful. Exposure of the MBP-atPEPC fusion protein to various amounts of thrombin resulted in a specific cleavage of the fusion protein and inactivation of the PEPC activity (data not shown). Due to these problems, we decided to perform a brief biochemical characterization of atPEPC with the purified MBP-atPEPC fusion protein.

**Physical and biochemical characterization of atPEPC.** The atPEPC from *S. solfataricus* displayed the highest activity at ca. 85°C (Table 2), which is close to the optimal temperature reported for the purified PEPC from *S. acidocaldarius* (90°C) (25). In addition, the enzyme activity was completely dependent on the presence of $Mg^{2+}$. The enzyme activity could not be reconstituted by the addition of $Mn^{2+}$ instead of $Mg^{2+}$ (Table 2). Moreover, *S. solfataricus* atPEPC displayed Michaelis-Menten kinetics with a $K_{m[PEP]}$ of 0.09 mM and a specific activity of 2.1 U/mg under the standard assay conditions (Table 2). These observations are in good agreement with the data reported for the purified *S. acidocaldarius* enzyme (25). Given the reported lack of allosteric regulation of the previously characterized archaeal PEPCs by the known effectors of BE-PEPC, we analyzed the effect of these effectors on the recombinant *S. solfataricus* atPEPC. Allosteric activators of *E. coli* PEPC (glucose 6-phosphate, fructose 1,6-bisphosphate, and acetyl coenzyme A) did not affect the activity of atPEPC (Table 2). This lack of allosteric activation could be explained by the absence, in atPEPC sequences, of counterparts of the regions of BE-PEPC that are involved in allosteric activation (see above). However, the addition of L-aspartate and L-malate to a final concentration of 5 mM to the standard enzyme assay resulted in inhibition of PEPC activity of 50 and 20%, respectively, which roughly corresponds to the characteristics of the *S. acidocaldarius* enzyme (25). The sensitivity of both *Sulfolobus* PEPC activities to L-aspartate is surprising given the absence of known allosteric domains. Conceivably, atPEPCs evolved a distinct mechanism of L-aspartate binding and regulation that is compatible with the low degree of conservation of the predicted L-aspartate-contacting residue in motif XI in the atPEPC family (Fig. 2). Purified PEPC from the archaeal methanogen *Methanothermus sociabilis* has been reported to be insensitive to L-aspartate-mediated repression, as well as to all other known PEPC effectors, including L-malate (26). The sequence of this particular atPEPC, which is currently unavailable, might provide clues as to the mechanisms of allosteric regulation in this archaeal protein family.

**Functional implications of atPEPC in archaeal central carbohydrate pathways.** The detection of a gene encoding PEPC activity sheds light on the organization of the central carbon metabolism in *Archaea*, especially in autotrophic species. Most sequenced archaeal genomes encode a partial TCA cycle (11) that is used as a source of intermediates (e.g., OAA, succinate, and α-ketoglutarate) for the biogenesis of amino acids and vitamins. For some methanogenic species, e.g., the *Methanococcales* and *Methanobacteriales*, a partial TCA cycle has been found to operate in a reductive direction from OAA to α-ketoglutarate (28). In contrast, methanogens from the order *Methanosarcinales* make use of a partial TCA cycle that operates in the oxidative direction from OAA to α-ketoglutarate (28). Therefore, it is obvious that OAA biosynthetic pathways play a central role in methanogenic archaeal species. The identification of the archaeal PEPC gene solves the problem of OAA formation for *Methanopyrus kandleri* (29). This finding also suggests that in methanogenic archaeal species and *A. fulgidus* atPEPC is an important link between gluconeogenesis and amino acid biosynthesis. This hypothesis is indirectly supported by the fact that citrate synthase is not encoded in the *Methanopyrus kandleri* and *Methanococcus jannaschii* genomes (Table 1). It should be noted, however, that, apart from the atPEPC gene, three archaeal methanogens and *A. fulgidus* also have genes encoding a heterodimeric pyruvate carboxylase (21,

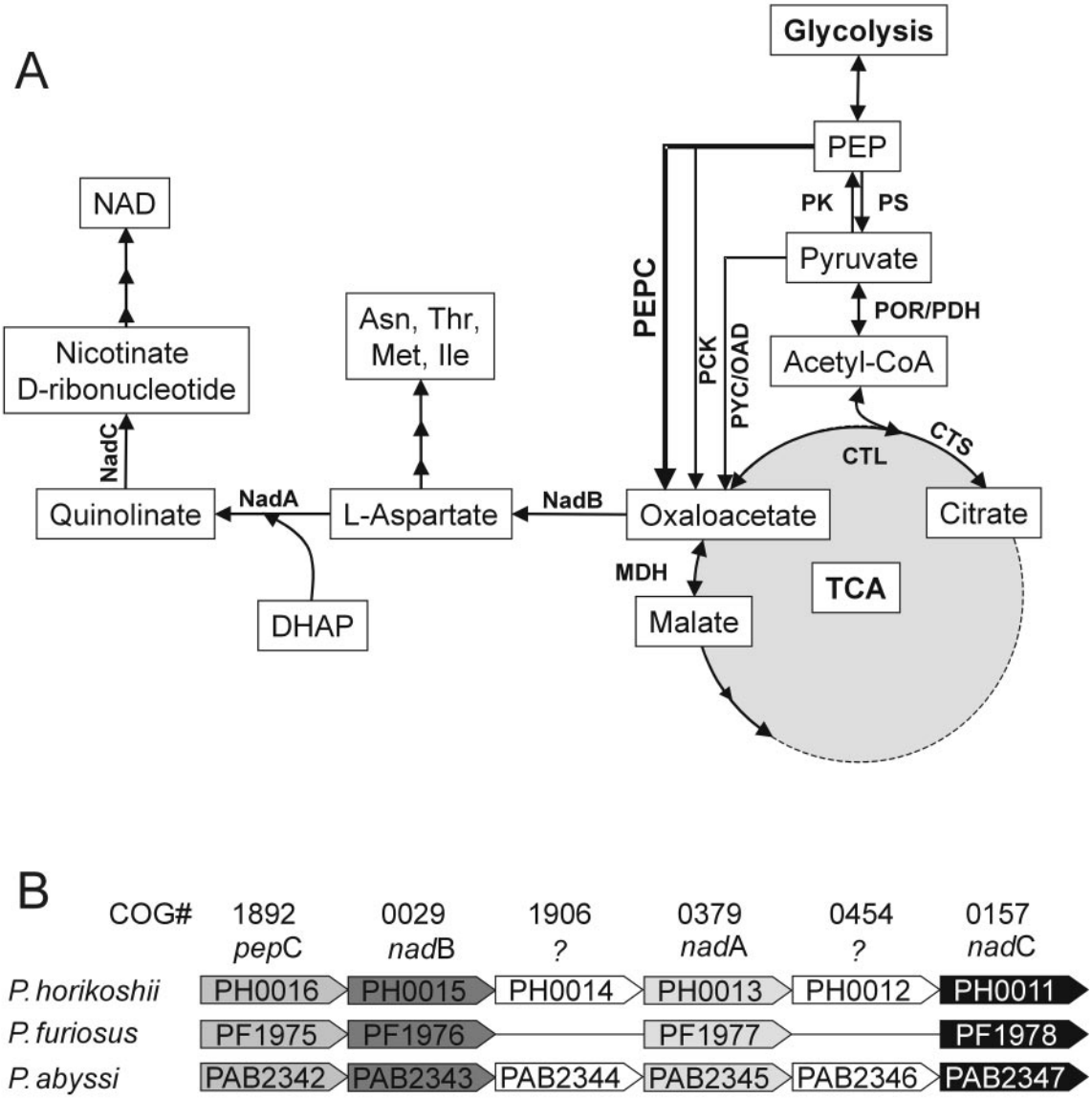


FIG. 4. (A). Overview of OAA-forming and -consuming pathways in *Archaea*. Abbreviations: PYC, pyruvate carboxylase; OAD, OAA carboxylase; PCK, PEP carboxykinase (ATP/GTP dependent); PK, pyruvate kinase; PS, PEP synthase; CTS, citrate synthase; CTL, ATP:citrate lyase; POR, pyruvate oxido-reductase; PDH, pyruvate dehydrogenase; NadA, quinolinate synthase; NadB, L-aspartate oxidase; NadC, nicotinate-nucleotide pyrophosphorylase (carboxylating); DHAP, dihydroxy-acetone-phosphate. (B) Conserved local gene context of atPEPC and aspartate/NAD biosynthesis genes pathway in *Pyrococcus*. Orthologous genes are shaded in the same grayscale, and corresponding COGs are depicted; genes encoding proteins with an unknown function are boxed white. Genes are not drawn to scale.

22) (COG5016/439), which is anticipated to be involved in OAA formation (Table 1 and Fig. 4A).

Whereas atPEPC is probably involved in OAA formation as part of a reductive TCA cycle in autotrophic *Archaea*, their most likely role in heterotrophic *Archaea* is solely anaplerotic. In the pyrococcal genomes, the atPEPC gene forms a putative operon, which additionally contains genes involved in L-aspartate and NAD biosynthesis in all three (Fig. 4B). Aspartate is an important precursor in the biosynthetic routes of several other amino acids, such as asparagine, methionine, leucine, and isoleucine (Fig. 4A), which makes aspartate a well-suited signal molecule for the demand for these amino acids in the cell. Not surprisingly, most PEPCs characterized to date, including the atPEPCs from heterotrophic *Archaea*, are subject to aspartate-mediated allosteric repression. An anaplerotic role for atPEPC in heterotrophic *Archaea* is suggested also by the recent whole-genome microarray analysis of *Pyrococcus furiosus* (27). The PEPC gene in this organism (PF1975) was 20-fold upregulated in a maltose-grown culture compared to a culture that was grown on peptides.

Identification, characterization, and comparison of enzymes from the archaeal variants of the Embden-Meyerhof, Entner-Doudoroff, and TCA pathways has revealed several instances of nonorthologous gene displacement, as well as some unique genes (for reviews, see references 33 and 11). The identified atPEPC, with its distinct properties, constitutes yet another

example of the versatility of the enzymes of the central carbon metabolic pathways in the archaeal domain.

## ACKNOWLEDGMENT

We thank Paul Riggs (New England Biolabs) for kindly providing plasmid pMAL-c2T.

## ADDENDUM

While the present study was in review, a report by Patel et al. (24a) was published that describes the purification and characterization of an archaeal PEPC of *Methanothermobacter thermoautotrophicus*. By determination of the N-terminal sequence, these authors identified the archaeal PEPC family that is also reported here. It should be noted that both the strategy of Patel et al. (a reverse genetics approach) and the strategy used here (a function prediction/verification via bioinformatics approach) resulted in the identification of the same archaeal PEPC family.

## REFERENCES

1. **Adachi, J., and M. Hasegawa.** 1992. MOLPHY: programs for molecular phylogenetics. I. PROTML: maximum-likelihood inference for protein phylogeny. *In* Computer science monographs, vol. 27. I.O.S. Mathematics, Tokyo, Japan.
2. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
3. **Aravind, L., and E. V. Koonin.** 1999. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J. Mol. Biol. **287:**1023–1040.
4. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:**248–254.
5. **Chollet, R., J. Vidal, and M. H. O'Leary.** 1996. Phospho*enol*pyruvate carboxylase: a ubiquitous, highly regulated enzyme in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **47**:273–298.
6. **Cuff, J. A., M. E. Clamp, A. S. Siddiqui, M. Finlay, and G. J. Barton.** 1998. JPred: a consensus secondary structure prediction server. Bioinformatics **14:**892–893.
7. **Davis, A. J., M. A. Perugini, B. J. Smith, J. D. Stewart, T. Ilg, A. N. Hodder, and E. Handman.** 2004. Properties of GDP-mannose pyrophosphorylase, a critical enzyme and drug target in *Leishmania mexicana*. J. Biol. Chem. **279:**12462–12468.
8. **Felsenstein, J.** 1996. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. Methods Enzymol. **266:**418–427.
9. **Fitch, W. M., and E. Margoliash.** 1967. Construction of phylogenetic trees. Science **155:**279–284.
10. **Fox, J. D., K. M. Routzahn, M. H. Bucher, and D. S. Waugh.** 2003. Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers. FEBS Lett. **537:**53–57.
11. **Huynen, M. A., T. Dandekar, and P. Bork.** 1999. Variation and evolution of the citric-acid cycle: a genomic perspective. Trends Microbiol. **7:**281–291.
12. **Jones, D. T., W. R. Taylor, and J. M. Thornton.** 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. **8:**275–282.
13. **Kai, Y., H. Matsumura, T. Inoue, K. Terada, Y. Nagara, T. Yoshinaga, A. Kihara, K. Tsumura, and K. Izui.** 1999. Three-dimensional structure of phospho*enol*pyruvate carboxylase: a proposed mechanism for allosteric inhibition. Proc. Natl. Acad. Sci. USA **96:**823–828.
14. **Kai, Y., H. Matsumura, and K. Izui.** 2003. Phospho*enol*pyruvate carboxylase: three-dimensional structure and molecular mechanisms. Arch. Biochem. Biophys. **414:**170–179.
15. **Kishino, H., T. Miyata, and M. Hasegawa.** 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. **31:**151–160.
16. **Koonin, E. V., A. R. Mushegian, and P. Bork.** 1996. Non-orthologous gene displacement. Trends Genet. **12:**334–336.
17. **Makarova, K. S., and E. V. Koonin.** 2003. Comparative genomics of archaea: how much have we learned in six years, and what's next? Genome Biol. **4:**115.
18. **Makarova, K. S., and E. V. Koonin.** 2003. Filling a gap in the central metabolism of archaea: prediction of a novel aconitase by comparative-genomic analysis. FEMS Microbiol. Lett. **227:**17–23.
19. **Matsumura, H., M. Terada, S. Shirakata, T. Inoue, T. Yoshinaga, K. Izui, and Y. Kai.** 1999. Plausible phospho*enol*pyruvate binding site revealed by 2.6 Å structure of $Mn^{2+}$-bound phospho*enol*pyruvate carboxylase from *Escherichia coli*. FEBS Lett. **458:**93–96.
20. **Matsumura, H., Y. Xie, S. Shirakata, T. Inoue, T. Yoshinaga, Y. Ueno, K. Izui, and Y. Kai.** 2002. Crystal structures of C4 form maize and quaternary complex of *E. coli* phospho*enol*pyruvate carboxylases. Structure **10:**1721–1730.
21. **Mukhopadhyay, B., V. J. Patel, and R. S. Wolfe.** 2000. A stable archaeal pyruvate carboxylase from the hyperthermophile *Methanococcus jannaschii*. Arch. Microbiol. **174:**406–414.
22. **Mukhopadhyay, B., E. Purwantini, C. L. Kreder, and R. S. Wolfe.** 2001. Oxaloacetate synthesis in the methanarchaeon *Methanosarcina barkeri*: pyruvate carboxylase genes and a putative *Escherichia coli*-type bifunctional biotin protein ligase gene (*bpl*/*birA*) exhibit a unique organization. J. Bacteriol. **183:**3804–3810.
23. **Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia.** 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. **247:**536–540.
24. **Notredame, C., D. G. Higgins, and J. Heringa.** 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. **302:**205–217.

24a. **Patel, H. M., J. L. Kraszewski, and B. Mukhopadhyay.** 2004. The phosphoenolpyruvate carboxylase from *Methanothermobacter thermautotrophicus* has a novel structure. J. Bacteriol. **186:**5129–5137.

25. **Sako, Y., K. Takai, T. Nishizaka, and Y. Ishida.** 1997. Biochemical relationship of phospho*enol*pyruvate carboxylases (PEPCs) from thermophilic archaea. FEMS Microbiology Lett. **153:**159–165.
26. **Sako, Y., K. Takai, A. Uchida, and Y. Ishida.** 1996. Purification and characterization of phospho*enol*pyruvate carboxylase from the hyperthermophilic archaeon *Methanothermus sociabilis*. FEBS Lett. **392:**148–152.
27. **Schut, G. J., S. D. Brehm, S. Datta, and M. W. Adams.** 2003. Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on carbohydrates or peptides. J. Bacteriol. **185:**3935–3947.
28. **Simpson, P. G., and W. B. Withman.** 1993. Anabolic pathways in methanogens, p. 445–472. *In* J. G. Ferry (ed.), Methanogenesis: ecology, physiology, biochemistry, and genetics. Chapman and Hall, New York, N.Y.
29. **Slesarev, A. I., K. V. Mezhevaya, K. S. Makarova, N. N. Polushin, O. V. Shcherbinina, V. V. Shakhova, G. I. Belova, L. Aravind, D. A. Natale, I. B. Rogozin, R. L. Tatusov, Y. I. Wolf, K. O. Stetter, A. G. Malykh, E. V. Koonin, and S. A. Kozyavkin.** 2002. The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. Proc. Natl. Acad. Sci. USA **99:**4644–4649.
30. **Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin, and D. A. Natale.** 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4:**41.
31. **Teraoka, H., K. Izui, and H. Katsuki.** 1972. Phospho*enol*pyruvate carboxylase of *Escherichia coli*: alteration of allosteric properties by photooxidation. Arch. Biochem. Biophys. **152:**821–827.
32. **Ting, I. P., and C. B. Osmond.** 1973. Multiple forms of plant phospho*enol*pyruvate carboxylase associated with different metabolic pathways. Plant Physiol. **51:**448–453.
33. **Verhees, C. H., S. W. Kengen, J. E. Tuininga, G. J. Schut, M. W. Adams, W. M. De Vos, and J. Van Der Oost.** 2003. The unique features of glycolytic pathways in *Archaea*. Biochem. J. **375**:231–246.
34. **Zeikus, J. G., G. Fuchs, W. Kenealy, and R. K. Thauer.** 1977. Oxidoreductases involved in cell carbon synthesis of *Methanobacterium thermoautotrophicum*. J. Bacteriol. **132:**604–613.