Reconstruction of the Central Carbohydrate Metabolism of *Thermoproteus tenax* by Use of Genomic and Biochemical Data

Bettina Siebers,1,* Britta Tjaden,1 Klaus Michalke,1 Christine Dörr,1 Hatim Ahmed,1 Melanie Zaparty,1 Paul Gordon,2 Christoph W. Sensen,2 Arne Zibat,3 Hans-Peter Klenk,3 Stephan C. Schuster,4 and Reinhard Hensel1

Department of Microbiology, Universität Duisburg-Essen, 45117 Essen,1 Epidauros Biotechnologie AG, 82347 Bernried,1 and Max-Planck Institut für Entwicklungsbiologie, 72076 Tübingen,4 Germany, and Sun Center of Excellence for Visual Genomics, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada T2N 4N12

Received 10 June 2003/Accepted 13 November 2003

The hyperthermophilic, facultatively heterotrophic crenarchaeum *Thermoproteus tenax* was analyzed using a low-coverage shotgun-sequencing approach. A total of 1.81 Mbp (representing 98.5% of the total genome), with an average gap size of 100 bp and 5.3-fold coverage, are reported, giving insights into the genome of *T. tenax*. Genome analysis and biochemical studies enabled us to reconstruct its central carbohydrate metabolism. *T. tenax* uses a variant of the reversible Embden-Meyerhof-Parnas (EMP) pathway and two different variants of the Entner-Doudoroff (ED) pathway (a nonphosphorylative variant and a semiphosphorylative variant) for carbohydrate catabolism. For the EMP pathway some new, unexpected enzymes were identified. The semi-phosphorylative ED pathway, hitherto supposed to be active only in halophiles, is found in *T. tenax*. No evidence for a functional pentose phosphate pathway, which is essential for the generation of pentoses and NADPH for anabolic purposes in bacteria and eucarya, is found in *T. tenax*. Most genes involved in the reversible citric acid cycle were identified, suggesting the presence of a functional oxidative cycle under heterotrophic growth conditions and a reductive cycle for CO₂ fixation under autotrophic growth conditions. Almost all genes necessary for glycogen and trehalose metabolism were identified in the *T. tenax* genome.

Archaeota were recognized as a distinct phylogenetic group more than 25 years ago (75). Their importance as the third major evolutionary line is well established, but our knowledge of their physiological capabilities remains limited. Central metabolic pathways within these organisms are far from fully understood (43, 47, 72).

*Thermoproteus tenax* was the first hyperthermophilic archaeon described (76). It is able to grow chemolithoautotrophically within these organisms are far from fully understood (43, 47, 72).

The hyperthermophilic, facultatively heterotrophic crenarchaeum *Thermoproteus tenax* was analyzed using a low-coverage shotgun-sequencing approach. A total of 1.81 Mbp (representing 98.5% of the total genome), with an average gap size of 100 bp and 5.3-fold coverage, are reported, giving insights into the genome of *T. tenax*. Genome analysis and biochemical studies enabled us to reconstruct its central carbohydrate metabolism. *T. tenax* uses a variant of the reversible Embden-Meyerhof-Parnas (EMP) pathway and two different variants of the Entner-Doudoroff (ED) pathway (a nonphosphorylative variant and a semiphosphorylative variant) for carbohydrate catabolism. For the EMP pathway some new, unexpected enzymes were identified. The semi-phosphorylative ED pathway, hitherto supposed to be active only in halophiles, is found in *T. tenax*. No evidence for a functional pentose phosphate pathway, which is essential for the generation of pentoses and NADPH for anabolic purposes in bacteria and eucarya, is found in *T. tenax*. Most genes involved in the reversible citric acid cycle were identified, suggesting the presence of a functional oxidative cycle under heterotrophic growth conditions and a reductive cycle for CO₂ fixation under autotrophic growth conditions. Almost all genes necessary for glycogen and trehalose metabolism were identified in the *T. tenax* genome.

The main difference between the *T. tenax* EMP variant and the classical version of this pathway concerns (i) the presence of a hexokinase (HK) with no notable regulatory properties (11), (ii) the replacement of the antagonistic enzyme couple ATP-dependent PFK and fructose bisphosphatase by a bidirectional PP₁-dependent enzyme (56), (iii) the presence of two different glyceraldehyde-3-phosphate (GAP) dehydrogenases (GAPDHs), differing in phosphate dependence, reversibility of the catalyzed reaction, and allosteric properties (4, 5, 21, 40) and (iv) a pyruvate kinase (PK) with reduced allosteric potential (48). Enzyme as well as transcript studies with *T. tenax* (late-log-phase cells) indicate that regulation takes place on the protein and gene levels. Under heterotrophic conditions,
the catabolic flux seems to be forced by the preferred expression of the PP-dependent phosphofructokinase (PP-PFK) and fructose-bisphosphate aldolase (FBPA), which form an operon in *T. tenax* (fbp-pfp operon) (34, 54, 56), and of the PK (48). A strong influence on the catabolic flux is also exerted by the allosterically regulated nonphosphorylating GAPDH (GAPN), which catalyze the unidirectional oxidation of GAP to 3-phospho-D-glycerate. In the presence of activators such as glucose-1-phosphate, fructose-6-phosphate, AMP, and ADP, i.e., under conditions which are characterized by the availability of storage carbohydrates (e.g., glycogen) and/or a low-energy charge of the cell, the catabolic carbon flux is most effective. In contrast, the classical, phosphorylating NADP+-dependent GAPDH and phosphoglycerate kinase (PGK) are of predominant importance for anabolism rather than for catabolism, as indicated by the enzymatic features of the NADP+-dependent GAPDH, by higher transcript amounts of both genes (gap and pgk), and by higher-level enzyme activities in autotrophically grown cells (5). Phosphoenolpyruvate synthetase (PEPS) activity was detected only in extracts of cells grown under autotrophic conditions (53); thus, the PEPS seems to exert a key function in driving the carbon flux into the anabolic direction.

Enzymatic studies with cell extracts (49, 55) and analyses of characteristic intermediates (55) identified the nonphosphorylative version of the ED pathway, which had been described originally for *S. solfataricus* (9) and *T. acidophilum* (7). In contrast to findings for the haloarchaeal semiphosphorylative pathway (with phosphorylation at the level of 2-keto-3-deoxygluconate [KDG]), in this version of the ED pathway phosphorylation takes place only at the level of glyceraldehyde generated by the KDG aldolase (KDGAA) so that the characteristic intermediate of the nonphosphorylative version. So far, only glucose dehydrogenase (GDH), the first enzyme of the nonphosphorylative ED pathway of *T. tenax*, has been purified and characterized (57).

Because of its essential function for providing pentoses and NADPH for anabolic purposes, the pentose phosphate pathway (PPP) has been assumed for all three domains of life. Genome analysis confirmed by biochemical studies indicates that glucose-6-phosphate dehydrogenase seems to be generally absent from methanarchaea and hyperthermophilic archaea (51, 52); therefore, the conventional PPP does not seem to be operative in these organisms. Also, in halophiles—although activity of glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase was detected in *Halococcus saccharolyticus* (23) and gene homologs of these enzymes were found in the genome of *Halobacterium* sp. strain NRC1—PPP is not involved in sugar catabolism, as indicated by 13C magnetic resonance studies of *H. saccharolyticus*. Possibly, the fragmentary PPP serves just for the synthesis of pentoses in haloarchaea.

The catabolic fate of pyruvate was monitored in *T. tenax* by fermentation analyses and enzyme measurements in crude extracts, which showed that the final oxidation of carbohydrates occurs via an oxidative citric acid cycle (CAC) linked to the glycolytic pathways by a pyruvate synthase (POR) (pyruvate:ferredoxin oxidoreductase) (49).

No information is available yet about the reaction sequence of CO₂ fixation in *T. tenax*. However, there is some evidence indicating that the CO₂ fixation occurs via the reductive CAC, as demonstrated for *T. neutrophilus*, a close relative of *T. tenax* (1). Most of the enzymes of the CAC work reversibly and could therefore be engaged in both directions. Only three counteracting enzyme pairs are thought to determine the oxidative or reductive direction of the cycle: (i) citrate synthase/citrate lyase, (ii) 2-oxoglutarate dehydrogenase/2-oxoglutarate synthase (2-oxoglutarate-ferredoxin oxidoreductase [KOR]), and (iii) succinate dehydrogenase/fumarate reductase. Experimental data (49) give no support for the presence of alternative CO₂ fixation mechanisms such as the Calvin cycle via ribulose-1,5-bisphosphate carboxylase or the reductive acetyl coenzyme A (acetyl-CoA) pathway via carbon monoxide dehydrogenase (Wood-Ljungdahl pathway).

Two additional carbohydrates—glycogen and trehalose—were identified in *T. tenax*, thus raising questions about their metabolism and function. Glycogen is a branched polymer of α-1,4-linked and α-1,6-linked glucosyl moieties and represents an osmotic inactive carbon storage compound in all three domains of life. Although the presence of glycogen in *T. tenax* and other archaea (members of *Thermoproteales* and *Sulfolobales*) was demonstrated previously (32), not much is known about glycogen metabolism in archaea. Trehalose, a glucose disaccharide [α-1,6-glucopyranosyl-(1→1)-α-1,6-glucopyranoside], is found in members of all three major phylogenetic domains. Trehalose plays an important role as a compatible solute, being involved in stress response to high-level osmolarity (*Escherichia coli*) and heat (*Saccharomyces cerevisiae*) (16, 39). Although trehalose has been identified in several archaea (e.g., *T. acidophilum*, *Methanothermobacter thermautotrophicus*, and *T. tenax*) (37, 39), the role of trehalose in archaea is still unknown. In *T. tenax*, trehalose was identified as the exclusive solute in concentrations of up to 0.3 μmol/mg of protein (37).

These insights into the physiological capabilities of *T. tenax* motivated us to study in more detail the biochemical and genetic mechanisms that select the various pathways, regulate the carbon flux through them, and govern the synthesis and breakdown of low- and high-molecular-mass carbohydrate compounds. For a better understanding of the central metabolism in *T. tenax*, as well as in archaea in general, we aimed at a representative genome sequence analysis of *T. tenax* to use the genetic information for reconstruction of the carbohydrate metabolism by assigning genes by sequence similarity and/or by the function of their recombinant gene products. So far, only four complete crenarchaeal genome sequences have been published: those for *Aeropyrum pernix* (26), *Pyrobaculum aerophilum* (13), and two species of *Sulfolobus* (25, 52). Here we report on the reconstruction of the *T. tenax*-specific EMP and ED pathways and the reversible CAC as well as on glycogen and trehalose metabolism.

### MATERIALS AND METHODS

**Strains and growth conditions.** Mass cultures of *T. tenax* Kral (DSM 2076) (12, 76) were grown under autotrophic conditions as described previously (5). *E. coli* strains DH5α (Invitrogen), BL21(DE3), and BL21-CodonPlus(DE3)-RIL (Stratagene) for use in cloning and expression studies were grown under standard conditions (46) following the instructions of the manufacturers.

**Library construction, shotgun sequencing, and primer walking.** Genomic DNA was prepared from *T. tenax* cells as described previously (48). For construction of a whole-genome shotgun library, genomic DNA from *T. tenax* was sheared with a HydroShear (GeneMachines, San Carlos, Calif.) to yield fragments with estimated sizes of 2.5 to 5.0 kbp. The fragmented DNA was separated
in a low-melting-point agarose gel, and the dominant fraction of the desired size range was isolated. The fragments were purified with a QIAquick gel extraction kit (Qiagen, Hilden, Germany) and cloned using a TOPO shotgun subcloning kit (Invitrogen Corp., Carlsbad, Calif.). Positive clones were identified by their ampicillin resistance in combination with their white color phenotype X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) screening. The isolated vectors served as templates in sequencing reactions using the primers M13uni(-43) (5′-aagcttgcaggcagctg-3′) and M13rev(-49) (5′-gagccgtaccccccag-3′). The sequence data were collected using an ABI Prism DNASystem 377 apparatus and an ABI Prism 3700 DNA analyzer and processed and assembled with the Finch TV software (University of Washington). Gaps were closed where needed by direct sequencing using chromosomal DNA as a template.

Automated analysis and annotation using the MAGPIE system. The genomic sequence was analyzed using Multipurpose Automated Genome Project Investigation Environment (MAGPIE) automated genome analysis and annotation software (15, 17). The function of open reading frames (ORFs) involved in metabolic pathways was verified manually, and gene ontology categories were attached to the functional assignments.

Cloning of genes and expression in E. coli. To prove the assignment of gene homologs in the T. tenax genome, the respective genes were cloned into a pET vector system via PCR mutagenesis. Expression of the gene products in E. coli BL21 (DE3) and BL21-CodonPlus(DE3)-RIL was performed following the instructions of the manufacturer (Stratagen). For expression, the glucose-6-phosphate isomerase (PGI) gene (pgi) of T. tenax was cloned into vector PET-11c through the use of the two new restriction sites (Ndel and BamHI) introduced by PCR mutagenesis with the following primer set: PGI-Ndel (5′-tccggattatactgctaagta-3′ [sense]) (the mutations are shown in boldface characters and the resulting restriction sites are underlined) and PGI-BamHI (5′-tggccgag gaatgcttgatc-3′ [antisense]). The resulting primer set—devised on the basis of N- and C-terminal partial sequence information—was designed to amplify the glycol phosphorylase (glgP) gene in pET-24a: GLGP-AesI (5′-gcccctccctatatgaatgagt-3′ [sense]) and GLGP-HindIIIrev (5′-agatcaacctgctcgccagctca-3′ [antisense]). Both enzymes were expressed in BL21-CodonPlus(DE3)-RIL.

Isolation of recombinant hyperthermophilic enzymes. Recombinant E. coli cells were suspended in buffer (e.g., 100 mM HEPES-KOH [pH 7.5] or 100 mM TRIS-HCl [pH 7.0 at 70 to 90°C, depending on the temperature for heat precipitation]) containing 7.5 mM dithiothreitol (DTT) or 10 mM β-mercaptoethanol and passed three times through a French pressure cell at 150 MPa. Cell debris and unbroken cells were removed by centrifugation (85,000 × g for 30 min at 4°C). The resulting crude extract was heat precipitated at 70 to 90°C for 20 to 30 min, depending on the stability of the respective enzyme. The fraction remaining after heat precipitation was centrifuged again (20,000 × g for 20 min at 4°C) and dialyzed overnight at 4°C. In general, after heat treatment only a few faint contaminant E. coli proteins were detectable by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie blue staining.

For the recombinant PGI, 0.5 g of E. coli cells was resuspended in 3 ml of 100 mM TRIS-HCl (pH 7.0 at 90°C) containing 10 mM β-mercaptoethanol. The heat precipitation was performed for 20 min at 90°C, and the protein was dialyzed overnight at 4°C with 1.6 liters of 100 mM TRIS-HCl (pH 7.0 at 55°C) containing 10 mM β-mercaptoethanol. For the recombinant glycol phosphorylase, 10 g of E. coli cells was resuspended in 30 ml of 100 mM HEPES-KOH (pH 7.5) containing 7.5 mM DTT and the heat precipitation procedure was performed for 30 min at 90°C. The extract was dialyzed overnight at 4°C with 2 liters of 50 mM HEPES-KOH (pH 7.5) containing 7.5 mM DTT, and the dialyzed fraction was applied on O-Sepharose Fast-Flow (Amersham Pharmacia Biotech) equilibrated in the same buffer and eluted with a linear salt gradient of 0 to 600 mM KCl. Fractions containing the homogenous enzyme solution were pooled, dialyzed, and concentrated.

Enzyme assays. PGI activity was determined in the anabolic direction (glucose-6-phosphate formation) at 50°C in a coupled assay with glucose-6-phosphate dehydrogenase (EC 1.1.1.49) from baker’s yeast as an auxiliary enzyme. The assay (total volume, 0.5 ml) was performed at 50°C in 100 mM TRIS-HCl (pH 7.0) in the presence of 1 mM NADP+, 1 U of glucose-6-phosphate dehydrogenase, 0.05 to 7 mM fructose-6-phosphate, and 4.5 µg of PGI after heat precipitation. Enzyme activity in the catabolic direction (fructose 6-phosphate formation) was assayed at 50°C in a coupled assay using ATP-dependent phosphofructokinase (EC 2.7.1.11; Bacillus steatorrhoeus), fructose-1,6-bisphosphate aldolase (EC 4.1.2.13; rabbit muscle), triosephosphate isomerase (TIM) (EC 5.3.1.1; rabbit muscle), and glyceraldehyde 3-phosphate dehydrogenase (EC 1.1.1.8; rabbit muscle) as auxiliary enzymes. The assay (total volume, 0.5 ml) was performed at 100°C in 100 mM TRIS-HCl (pH 7.0) in the presence of 1 mM NADH, 3 mM ATP, 5 mM MgCl2, 2 U of phosphofructokinase, 1 U of fructose-1,6-bisphosphate aldolase, 50 U of TIM, 9 U of glyceraldehde 3-phosphate dehydrogenase, 0.1 to 14 mM glucose 6-phosphate, and 0.8 µg of recombinant PGI after heat precipitation.

RESULTS AND DISCUSSION

Genome sequence features. The genome assembled used for the annotation of the genes described in this paper consists of 19,850 sequence reads assembled in 53 contigs more than 2 kb in length; the largest contig is 110 kb in length. Contigs smaller than 2 kb were excluded from the analysis. The total length of the contigs adds up to 1.81 Mbp, which represents about 98.5% of the estimated genome size of 1.84 Mbp. The average gap size is estimated to be about 100 bp. With sequence coverage at 5.3-fold, the sequence quality of the trimmed contigs is approximately 15 errors in 10,000 bp. The automated analysis tool identified and annotated 3,913 ORFs, which were then surveyed for the presence of genes related to central carbohydrate metabolism. Given an average G+C content of the genome is calculated to be 54%. The individual contigs were subjected to an automated analysis with MAGPIE software. The ORFs were identified on the basis of stop codons and their upstream start codons. An ORF size of 100 amino acids was chosen as the minimum cutoff level. The automated analysis tool identified and annotated 3,913 ORFs, which were then surveyed for the presence of genes related to central carbohydrate metabolism. Given an average G+C level of 54%, this leads to an overprediction of ORFs; however, at least 2,000 to 2,500 of these should be coding ORFs (given the number of coding regions in other organisms of the same size).

The central carbohydrate metabolism of T. tenax. (i) Variant of the EMP pathway. (a) Assignment of genes involved in the reversible EMP variant. The following enzymes (and their coding genes) essential for a functional reversible EMP pathway have been identified and characterized to date: PP-PFK (56), FBPA (34, 54), GAPDH (5), PGK (5), TIM (H. Walden, G. L. Taylor, E. Lorentzen, E. Pohl, H. Lilie, A. Schramm, T. Knura, K. Stubbe, B. Tjaden, and R. Hensel, submitted for publication), PK (48), and PEPS (63; B. Tjaden, C. Dörre, B. Siebers, and R. Hensel, unpublished data). As an additional constituent of the catabolic version of the pathway, the non-phosphorylating GAPN (4, 5, 21), and its coding gene were identified.

Genes still missing from the T. tenax genome for a functional EMP pathway were assigned (Table 1; Fig. 1). In the case of HK (11), enolase (ENO), PGI, and phosphoglycerate mutase (PGAM), the sequence similarity of the encoding genes allowed an unequivocal assignment. The HK was originally assigned as glucokinase, and only the characterization of the enzyme after expression of the respective gene in E. coli revealed that it is a true HK, catalyzing the phosphorylation of different sugars such as glucose, fructose, and mannose (11).
Glucose

HK

PGI

FBP/IMP

Fru 6-P

PGP

PP\(_7\)-PFK

TIM

DHAP

GAP

1,3-DPG

GAPN

GAPDH

PGK

3-PG

PGAM

2-PG

PEP

PEPS

PK

PPDK

Pyruvate

The PGI showed similarity only to an annotated homolog of *P. aerophilum*, and no similarity to the recently described unique PGIs of the cupin family identified in archaea (e.g., *P. furiosus*) (19, 71) and bacteria or to the classical PGIs found in bacteria and eucarya was observed. The *pgi* gene (897 bp) of *T. tenax* encodes a protein of 298 amino acids with a calculated molecular mass of 33.2 kDa. The heterologous expression of PGI resulted in an enzyme with an apparent molecular mass of 33 kDa (as detected by SDS-PAGE) which catalyzes the reversible glucose 6-phosphate/fructose 6-phosphate conversion. The enzyme follows Michaelis-Menten kinetics for glucose 6-phosphate and fructose 6-phosphate were 2 mM and 0.17 mM, respectively; the \(V_{\text{max}}\) value for the glucose 6-phosphate/fructose 6-phosphate isomerization was determined to be 170 U/mg of protein and for the reverse anabolic reaction was determined to be 44.5 U/mg of protein (assay temperature, 50°C). This third, new type of PGI differs significantly at the sequence level from members of the cupin family and classical PGIs, and characteristic sequence motifs identified for members of the two other families are absent (19, 71). BlastX searches revealed homologs of this new PGI family in archaea (*P. aerophilum*, *T. acidophilum*, *T. volcanium*, *A. pernix*, *Ferroplasma acidarmanus*, *Sulfolobus tokodaii*, and *S. solfataricus*) and extreme thermophilic bacteria (*Anaerocellum thermophilum* and *Aquifex aeolicus*). The gene coding for the putative PGAM (partial sequence information) showed high similarity to those coding for the recently characterized enzymes of *Methanococcus jannaschii* (AAAB99632) (18, 68) and *P. furiosus* (AAL82083) (68).

*T. tenax* does not contain genetic information for glucose-6-phosphatase, suggesting either that glucose-6-phosphatase is absent or that *T. tenax* possesses a novel enzyme type that has not yet been recognized as such. In addition, gene homologs for ADP-dependent glucokinase or ADP-dependent phosphofructokinases present in other anaerobic, hyperthermophilic archaea (e.g., *P. furiosus*, *A. fulgidus*, and *Thermococcus* species) and for ATP-dependent PFK (Desulfurococcus species) are absent from the *T. tenax* genome, underlining the central function of the reversible nonallosteric PP\(_7\)-PFK.

Surprisingly, three ORFs in the *T. tenax* genome showed convincing similarities to genes encoding enzymes with functions obviously equivalent to those of the assigned EMP enzymes fructose-1,6-bisphosphatase (FBP), ferredoxin-dependent GAP oxidoreductase (GAPOR), and pyruvate, phosphate dikinase (PPDK) (partial sequence information). The identified FBP homolog is a member of the recently described archaeal FBP (type IV) and related enzymes of the inositol-monophosphatase family (COG 0483) and shows 28% identity to the characterized FBP of *P. furiosus* (70) and 29% identity to the inositol-1-monophosphatase (IMP) of *M. jannaschii* (61). More recently a type V FBP (COG 1980) was detected in *Thermococcus kodakaraensis*, and homologs were identified in almost all archaeal genomes (42, 72). A type V FBP which is supposed to be the general archaeal gluconeogenetic FBP (72) is absent from *T. tenax*. Thus, biochemical studies of the *T. tenax* enzyme have to be awaited to elucidate a possible gluconeogenetic role in parallel to that of the bifunctional nonallosteric PP\(_7\)-PFK. Ferredoxin-dependent GAPOR—like GAPN of *T. tenax*—catalyzes the nonphosphorylating, irreversible conversion of glyceraldehyde 3-phosphate to 3-phosphoglycerate and thus also fulfills a catabolic function. The conditions under which the expression of GAPOR is favored or discriminated against are yet unknown. The complete sequence information for the PPDK gene was obtained by the identification (via Southern hybridization) and sequencing of a genomic clone. The assignment of the PPDK homolog was confirmed by the functional properties of the encoded enzyme (63; Tjaden et al., unpublished data). The recombinant enzyme expressed in *E. coli* catalyzes the interconversion of PEP and pyruvate according to the following equation: \(\text{ATP} + \text{P} + \text{pyruvate} \leftrightarrow \text{AMP} + \text{PP} + \text{PEP}\). The reversibility of the PPDK reaction implies that the direction of conversion depends strongly on the ratio of the substrates and cosubstrates on both sides of the equation; therefore, the enzyme fulfills the functions of both the anabolic PEPs and the catabolic PK.

In summary, the genome data indicate that the variant of the EMP pathway in *T. tenax* is much more complex than previously suspected. In contrast to results seen with the classical pathway, no control point is present in the phosphorylating part of the EMP variant, as indicated by the presence of a HK with apparently reduced allosteric potential and a reversible nonallosteric PP\(_7\)-dependent PFK. The major control point is shifted to the middle of the pathway at the level of GAP, where three different enzymes (GAPN, GAPDH, and GAPOR) are involved in the regulation of the carbohydrate flux. In addition,
a second control point seems to emerge at the level of pyruvate/PEP conversion, which is catalyzed again by three different enzymes (PK, PEPS, and PPDK) (48, 63).

Other modifications of the EMP pathway are found in the anaerobic species *P. furiosus* (27, 28, 31, 66, 69), *Thermococcus* species (31, 44, 50), *D. amylolecticia* (20), and *A. fulgidus* (33). The glycolysis of *D. amylolecticia* differs from that of the classical version of the EMP pathway by the presence of a ferredoxin-dependent GAPOR replacing the “conventional” phosphate- and pyridine nucleotide-dependent GAPDH. A more extended variation is seen with *P. furiosus* and *Thermococcus* species (as well as with *A. fulgidus*), which additionally replace the classical ATP-dependent HK and phosphofructokinase with nonallosteric ADP-dependent glucokinases and phosphofructokinases. In addition, ADP-dependent phosphofructokinase activity was demonstrated in crude extracts from members of the orders *Methanococcales* (e.g., *Methanococcus maripaludis*) and *Methanosarcinales* (e.g., *Methanosarcina mazei*) (73) and a novel bifunctional ADP-dependent glucokinase/phosphofructokinase is described for *M. jannaschii* (45, 73). The absence of allosteric sugar kinases and the presence of GAPN and GAPOR enzyme activities in several archaea, as well as the identification of GAPN and GAPOR gene homologs—in addition to GAPDH—in many archaeal genomes (e.g., those of *Pyrococcus* spp.) (72), indicates that among all the different modifications of the EMP pathway in archaea, the regulation at the level of GAP might represent a unique feature of central carbon metabolism.

(b) **Organization of the EMP genes.** Analyses of gene organization and transcription studies reveal that several EMP genes are organized in operons, suggesting specific functional relationships between the genes or their products (Table 1) (see Fig. 3). Thus, as previously shown (54), the *pfp* gene coding for the PP↓-dependent PFK, which catalyzes the irreversible phosphorylation of fructose 6-phosphate, forms an operon with the *fba* gene encoding FBA (*fba-pfp* operon). The coordinated transcription of these genes in one operon seems meaningful for *T. tenax*, because both enzymes catalyze not only consecutive but also reversible reactions and are therefore of equivalent levels of importance for both directions of the pathway. Also, the phosphorylating NADP↓-dependent GAPDH gene (*gap*) and the PGK gene (*pgk*) are clustered in an operon (*pgk-gap* operon) (5). In both operons, the two genes overlap by 4 bp (the coding regions overlap by 1 bp) and the formation of bicistronic as well as monocistronic mRNAs was demonstrated. Northern analyses reveal that the HK gene (*hsk*) is part of an operon (*orfX-hsk* operon) (11) and that, again, both genes overlap by 4 bp. The ORF located upstream (*orfX*) codes for a protein of unknown function, and only one homolog was identified in *P. aerophilum* (AE009933) (11).

Remarkably, as shown by Northern analysis the TIM gene (*tpi*) forms an operon with the aconitase gene (*acn*), a constituent of the reversible CAC in *T. tenax*. The *acn* and *tpi* genes are separated by 40 bp. The *tpi* gene overlaps by 15 bp with a gene homolog coding for a putative nicotinate phosphoribosyl transferase (*pncB* gene; AJ515539), which is oriented in the opposite direction. Further downstream of the *pncB* gene (151 bp) is the gene coding for citrate synthase 1 (*cis1*) (63). This clustering of the *tpi* gene with CAC genes is rather unusual. In most archaea and bacteria analyzed so far, the *tpi* gene is adjacent to other EMP genes such as the *gap*, *pgk*, *gpmA*, and *eno* genes (63). The specific linkage of the *tpi* and *acn* genes in *T. tenax* may reflect the preferred functional relationship of

---

### TABLE 1. Identified genes coding for the enzymes of the variant of the EMP pathway in *T. tenax*

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>EC no.</th>
<th>COG no.</th>
<th>Gene</th>
<th>Accession no.</th>
<th>Evidence value and best hit</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hexokinase (HK)</td>
<td>2.7.1.1</td>
<td>1940</td>
<td>hsk</td>
<td>AJ510140</td>
<td>3e-99 <em>P. aerophilum</em>, glucokinase, (ROK family)</td>
<td>11</td>
</tr>
<tr>
<td>Glucose-6-phosphate isomerase (PGI)</td>
<td>5.3.1.9</td>
<td>0166</td>
<td>pgi</td>
<td>AJ622172</td>
<td>le-93 <em>P. aerophilum</em></td>
<td>56</td>
</tr>
<tr>
<td>PP↓-dependent phosphofructokinase (PF↓-PFK)</td>
<td>2.7.1.90</td>
<td>0205</td>
<td>pfk</td>
<td>Y14655</td>
<td>0.0 <em>T. tenax</em></td>
<td>63</td>
</tr>
<tr>
<td>Fructose-1,6-bisphosphatase type IV (FBP)</td>
<td>3.1.3.11/1.3.25</td>
<td>0483</td>
<td>fbpA/subB</td>
<td>le-63 <em>P. aerophilum</em>, extragenic suppressor protein SubB homolog</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Fructose-bisphosphate aldolase (FBPA)</td>
<td>4.1.2.13</td>
<td>1830</td>
<td>fba</td>
<td>AJ310483</td>
<td>1.0e-148 <em>T. tenax</em></td>
<td>54</td>
</tr>
<tr>
<td>Triosephosphate isomerase (TIM)</td>
<td>5.3.1.1</td>
<td>0149</td>
<td>tpi</td>
<td>AJ012066/AJ515539</td>
<td>1.0e-120 <em>T. tenax</em></td>
<td>74</td>
</tr>
<tr>
<td>Glyceraldehyde-3-phosphate dehydrogenase (nonphosphorylating) (GAPN)</td>
<td>1.2.1.9</td>
<td>1012</td>
<td>gapN</td>
<td>Y10625</td>
<td>0.0 <em>T. tenax</em></td>
<td>4, 5, 21, 40</td>
</tr>
<tr>
<td>Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) (GAPDH)</td>
<td>1.2.1.13</td>
<td>0057</td>
<td>gap</td>
<td>Y10126</td>
<td>1.0e-135 <em>P. aerophilum</em></td>
<td>5, 21</td>
</tr>
<tr>
<td>Glyceraldehyde-3-phosphate oxidoreductase (ferredoxin dependent) (GAPOR)</td>
<td>2.7.2.3</td>
<td>0126</td>
<td>pgk</td>
<td>Y10126</td>
<td>1.0e-124 <em>P. aerophilum</em></td>
<td>5</td>
</tr>
<tr>
<td>Phosphoglycerate mutase (PGAM)</td>
<td>5.4.2.1</td>
<td>3635</td>
<td>gpmA</td>
<td>AJ6212333</td>
<td>1e-80 <em>P. aerophilum</em>, phosphonopyruvate decarboxylase</td>
<td>56</td>
</tr>
<tr>
<td>Enolase (ENO)</td>
<td>4.2.1.11</td>
<td>0148</td>
<td>eno</td>
<td>AJ621325</td>
<td>0.0 <em>P. aerophilum</em>, aldehyde ferredoxin oxidoreductase</td>
<td>48</td>
</tr>
<tr>
<td>Pyruvate kinase (PK)</td>
<td>2.7.1.40</td>
<td>0469</td>
<td>pyk</td>
<td>AF965890</td>
<td>0.0 <em>T. tenax</em></td>
<td>63, Tjaden et al., unpublished</td>
</tr>
<tr>
<td>Phosphoenolpyruvate synthetase (PEPS)</td>
<td>2.7.9.2</td>
<td>0574</td>
<td>pps</td>
<td>AJ515537</td>
<td>0.0 <em>P. aerophilum</em></td>
<td>63, Tjaden et al., unpublished</td>
</tr>
<tr>
<td>Pyruvate, phosphate dikinase (PPDK)</td>
<td>2.7.9.1</td>
<td>0574</td>
<td>ppdk</td>
<td>AJ515538</td>
<td>0.0 <em>P. aerophilum</em></td>
<td>63, Tjaden et al., unpublished</td>
</tr>
</tbody>
</table>

---

*a* Enzymes exclusively catalyzing the catabolic pathway are underlined; those operating exclusively in the anaerobic direction are shown in bold. For enzymes that are biochemically characterized, the respective references are given. The abbreviations in parenthesis correspond to the respective enzymes shown in Fig. 1.


*c* Evidence values represent the best Blastp hits and respective species; for homologs with functions deviating from those of the assigned *T. tenax* enzymes, the respective annotations are given.
both reversible pathways in this organism and hints at their strong coordination under anabolic and catabolic conditions.

(ii) Variants of the Entner-Doudoroff (ED) pathway. (a) Assignment of genes involved in the non- and semiphosphorylative ED pathways. For reconstruction of the nonphosphorylative ED pathway of T. tenax, we screened for genes encoding the enzymes GDH, gluconolactonase, gluconate dehydratase (GAD), KDGK, aldehyde oxidoreductase/dehydrogenase, and glycerate kinase. Homologs of all these genes—albeit with exception of the glycerate kinase gene—are found in the T. tenax genome. The central intermediate KDG of both variants is boxed. Closed arrows indicate the reactions of the nonphosphorylative version of the ED pathway; open arrows indicate the reactions of the semiphosphorylative version. Grey-shaded arrows mark reactions involved in both versions. For abbreviations of enzyme names, see Table 2.

The assignment of genes involved in the non- and semiphosphorylative ED pathways was based only on sequence similarity to the putative genes of S. solfataricus, T. acidophilum, P. furiosis, and Thermotoga maritima (29, 26, 33, and 34% identity, respectively, by a Blast search). The gene encoding the enzyme which catalyzes the oxidation of glyceraldehyde to glycerate could not be assigned decisively. In the T. tenax genome, three ORFs (for candidate 3, only partial C- and N-terminal information is available) were found which show convincing similarities to genes encoding aldehyde-ferredoxin oxidoreductases; the sequences of two ORFs show high-level similarity to sequences annotated as genes coding for aldehyde dehydrogenases or for related enzymes which are—like GAPN—members of the aldehyde dehydrogenase superfamily. Since cell extracts of T. tenax were reported to catalyze the oxidation of glyceraldehyde with benzylviologen only but not with NADP$^+$ or NAD$^+$ as electron acceptors (49), we assume that ferredoxin-dependent oxidoreductases is responsible for oxidation of glyceraldehyde. A gene homolog for ferredoxin was identified in the T. tenax genome.

Surprisingly, an ORF with high-level similarity to genes encoding KDG kinases (KDGKs) and fructokinases was identified in the T. tenax genome. The close neighborhood to the kdgA gene suggests an involvement in the ED pathway (see below). The KDGK gene was expressed heterologously in E. coli and the corresponding enzyme activity has been confirmed, indicating that the semiphosphorylative ED pathway is also active in T. tenax (Ahmed et al., unpublished data). This finding is rather unexpected, since the semiphosphorylative ED pathway has been assumed to be characteristic of haloarchaea. The localization of a gene homolog coding for glucan-1,4-α-glucosidase (GAA) has been assumed to be characteristic of haloarchaea in T. tenax (see below). A homolog of the glucose-6-phosphate dehydrogenase gene of the kind typical of the classical ED pathway was not found in the T. tenax genome.

Thus, the semiphosphorylative ED pathway seems to be present in this hyperthermophile probably in addition to the nonphosphorylative ED variant, and the KD(P)GA is the key activity in both variants. The organization of ED genes suggests a preference for the function of the ED variants in the hydrolytic degradation of polysaccharides (e.g., glycogen) in T. tenax (see below). A homolog of the glucose-6-phosphate dehydrogenase gene is found in the T. tenax genome.

The absence of the gluconolactonase homolog in T. tenax was based only on sequence similarity to the putative genes of S. solfataricus, T. acidophilum, P. furiosis, and Thermotoga maritima (29, 26, 33, and 34% identity, respectively, by a Blast search). The gene encoding the enzyme which catalyzes the oxidation of glyceraldehyde to glycerate could not be assigned decisively. In the T. tenax genome, three ORFs (for candidate 3, only partial C- and N-terminal information is available) were found which show convincing similarities to genes encoding aldehyde-ferredoxin oxidoreductases; the sequences of two ORFs show high-level similarity to sequences annotated as genes coding for aldehyde dehydrogenases or for related enzymes which are—like GAPN—members of the aldehyde dehydrogenase superfamily. Since cell extracts of T. tenax were reported to catalyze the oxidation of glyceraldehyde with benzylviologen only but not with NADP$^+$ or NAD$^+$ as electron acceptors (49), we assume that ferredoxin-dependent oxidoreductases is responsible for oxidation of glyceraldehyde. A gene homolog for ferredoxin was identified in the T. tenax genome.

(b) Organization of the ED genes. Analysis of the organization of ED genes (Table 2; Fig. 2) revealed the presence of a gene cluster which comprises the genes coding for GAD (gad), KD(P)GA (kdgA), KDGK (kdgK), and GAA (gaa). The kdgA and kdgK genes as well as the kdgA and gaa genes overlap by 4 bp, giving good evidence that the three genes form an operon
(Fig. 3B). The gene encoding the GAD is localized 68 bp downstream of the putative kdgA-kdgK-gaa operon and is oriented in the opposite direction (Fig. 3A).

(iii) Pentose metabolism. Assignment of genes involved in pentose metabolism. *T. tenax* does not possess the complete gene set for oxidative degradation of carbohydrates via the PPP (Table 3). As seen with *S. solfataricus* (52), only the gene homologs for ribosephosphate isomerase and transketolase (partial sequence information for the N- and C-terminal sections) were found to date. Genes coding for the classic glucose-6-phosphate dehydrogenase, decarboxylating phosphogluconate dehydrogenase, ribulosephosphate-3-epimerase, and transaldolase seem to be absent. These findings might indicate that hyperthermophilic archaea either use different, nonhomologous enzymes (which show no sequence similarity to known counterparts) in the PPP or utilize a yet unknown, analogous pathway for carbohydrate catabolism. Thus, the biosynthesis of pentoses in hyperthermophilic archaea remains unclear. The presence of a nonoxidative PPP (as proposed for ribose biosynthesis in *Methanococcus maripaludis* (67)) seems to be improbable, since a gene homolog for the common ribosephosphate-3-epimerase was not found in the *T. tenax* genome. However, two ORFs were detected in the *T. tenax* genome which showed similarity to genes coding for well-known enzymes of pentose metabolism such as ribokinase (sugar kinase of the ribokinase [pfkB] family) (ATP + d-ribose ↔ ADP + d-ribose 5-phosphate) and deoxyribose-phosphate aldolase (2-deoxy-d-ribose 5-phosphate ↔ d-glyceraldehyde 3-phosphate + acetaldehyde).

(iv) The reversible CAC. (a) Assignment of genes involved in the oxidative and reductive CAC. As shown in Table 4, not all enzymes necessary for the oxidative and reductive CAC were identified in the *T. tenax* genome. For some enzymes, not all expected subunits could be found or a confident assignment is not possible because of intermediate levels of similarity to different enzymes. However, most of the enzymes could be assigned due to the high sequence similarity of their encoding genes to sequences of gene homologs (such as those encoding citrate synthase, aconitase, isocitrate dehydrogenase, succinyl-CoA synthetase, fumarase, and malate dehydrogenase) in other genomes. All CAC genes, with the sole exception of the citrate synthase-encoding gene, of which two copies (cis1 and cis2) were found, occurred only once in the genome. The *cis1* gene was originally identified in the genomic clone harboring the genes encoding TIM and aconitase. The enzyme was expressed in *E. coli* and showed typical citrate synthase activity (63). The expression of *cis2* is planned for the near future. Detailed enzymatic as well as transcription studies of both genes have to be completed before the physiological rationale behind the gene doubling can be resolved.

In addition to the gene encoding isocitrate dehydrogenase, a second gene homolog was identified which showed high similarity to genes encoding isocitrate dehydrogenases as well as 3-isopropylmalate dehydrogenases (*leuB*; AJ621334). However, due to its localization downstream of a putative *leuA* gene homolog, which codes for 2-isopropylmalate synthase (AJ621335), we assume that this homolog represents the *leuB* gene, which is involved in the leucine biosynthesis pathway (22). For the reversible fumarase (fumarate hydratase), three different ORFs were found in the *T. tenax* genome: one gene homolog (partial sequence information) with high similarity to genes encoding class II enzymes and two genes which show similarity either to the N-terminal and C-terminal domains, respectively, of genes encoding bacterial class I fumarases or alternatively to those encoding subunits α and β of tartrate dehydratases. Although both fumarase homologs have been annotated for many archaeal genomes, to our knowledge only the class II fumarase from *S. solfataricus* has been characterized (8).

The identification of the genes encoding enzymes determining the direction of the cycle—with the exception the gene encoding citrate synthase—reveals some problems. (i) For the citrate lyase, only a gene homolog for the β chain [citryl-CoA lyase subunit; (3S)-citryl-CoA ↔ acetyl-CoA + oxaloacetate]...
was found but homologs for subunits α [citrate-CoA transferase subunit; acetyl-CoA + citrate ↔ acetate + (3S)-citryl-CoA] and γ (acyl carrier protein subunit) are missing [Lacunostoc mesenteroides (αβγ) subunit composition] (2). To our knowledge nothing is known so far about the reaction mechanism of archaeal citrate lyases, although ATP citrate lyase activity was demonstrated in cell extracts of T. netrophilus (1). The gene coding for the β subunit seems to form an operon with three gene homologs coding for a putative transport protein (pip; AJ621305), putative acetyl-CoA synthetases (acetate-CoA ligase gene and acs; AJ621304), and acetyl-CoA transferases or carnitine dehydratases (act; AJ621303). Thus, acetate-CoA ligase (EC 6.2.1.1; ATP + acetate + CoA ↔ AMP + Pi + acetyl-CoA) might be responsible for the gener-
activation of acetyl-CoA, which is considered to be a free analog of the acetylated acyl carrier protein (γ subunit) (2) and the putative acetyl-CoA transferase might represent the citrate-CoA transferase subunit. However, for a confidential assignment of citrate lyase experimental analyses have to be awaited. (ii) For the catabolic enzyme 2-oxoglutarate dehydrogenase, only the E3 component (dihydrolipoamide dehydrogenase; two candidates) of the multienzyme complex consisting of E1, dihydrolipoamide-S-succinyltransferase (E2), and E3 was found, which can be shared by different 2-oxo acid dehydrogenases and the glycine cleavage system (22, 74). (iii) KOR, which is thought to catalyze the reaction in the anabolic direction, could not be assigned unequivocally because of the high level of similarity between the different ferredoxin-dependent oxidoreductases. Two clusters of gene homologs were identified with similarity to genes encoding POR
pyruvate:ferredoxin oxidoreductase) and related 2-oxoacid syntheses (2-oxoacid:ferredoxin oxidoreductases [OOR]). Both are organized in an operon which comprises genes encoding the α, β, γ, and δ subunits and thus—because of their four-subunit composition—might represent either POR or KOR (22, 35, 62). In addition, four additional gene clusters were identified which encode homologs for 2-oxoacid synthases (accession no. AJ621306, AJ621307, AJ621313, AJ621314, AJ621318, AJ621319, AJ621320, AJ621352, and AJ621353).

(iv) For the succinate dehydrogenase (catabolic) and fumarate reductase (anaerobic) in a gene cluster with homologs coding for the putative succinate dehydrogenase (candidate 1) is organized in an operon which encodes all four subunits (α, β, γ, and δ), whereas the fumarate reductase (candidate 2) is organized in a gene cluster with homologs coding for the flavoprotein subunit (chain α), the iron-sulfur protein (chain β; homology only to the N-terminal part), and a zinc-dependent class III alcohol dehydrogenase (AJ621280). The gene displacement in noncatalytic subunits of fumarate reductases has been previously described (22); thus, whereas the membrane-bound enzymes are composed of four or three subunits, the cytoplasmic enzyme exhibits only one type of subunit.

In summary, all these findings suggest that the CAC in T. tenax is operative in the reductive and oxidative direction. The only missing enzyme is 2-oxoglutarate dehydrogenase. However, measurements in crude extracts demonstrate 2-oxoglutarate:benzyliogen oxidoreductase activity (49) and more recent reports about the POR in Clostridium thermoaceticum (14) demonstrate that the enzyme is active with ferredoxin as a cosubstrate in the oxidative as well as in the reductive reaction. Thus, 2-oxoglutarate oxidoreductase might be involved in both the oxidative and the reductive CAC in T. tenax.

No convincing candidates for the two key enzymes of the glyoxylate cycle—isocitrate lyase and malate synthase—were identified, suggesting that a functional glyoxylate cycle is absent from T. tenax. Surprisingly, putative gene homologs coding for the three subunits of the carbon monoxide dehydrogenase (medium, small, and large chains; AJ621345, AJ621344, and AJ621343, respectively) were identified in the T. tenax genome. All three subunits overlap by 4 bp and thus obviously form an operon. However, none of the other genes for a functional Wood-Ljungdahl pathway (e.g., from formylmethanofuran to methyl-tetrahydromethanopterin) were found, suggesting that a reductive acetyl-CoA pathway, which is described for several archaea (e.g., M.thermautotrophicus, Methanopyrus kandleri and in autotrophic species of Archaeoglobus [47]), is absent from T. tenax. However, the carbon monoxide dehydrogenase might be involved in coupling the POR reaction to CO oxidation via ferredoxin, as described for C. thermoaceticum (14).

(b) Organization of the CAC genes. Obviously, several CAC genes are organized in operons or operon-like structures (Fig. 3; Table 4). Seven operons have been identified so far. The gene coding for the β subunit of citrate lyase (citE) seems to form an operon with the upstream-located gene homologs coding for a putative transport protein (ptp), acetyl-CoA synthetase (acs), and acetyl-CoA transferases (act) or carnitine dehydratases. The ptp and acs as well as the act and citE genes overlap by 4 bp and the act and acs genes overlap by 17 bp, forming the ptp-acs-act-citE operon and suggesting a functional connection of the genes. Further, the genes coding for the reversible enzymes TIM (tpi) and aconitase (acn) (see above) are linked in an operon (acn-tpi) (as confirmed by Northern analysis) (63), indicating a strong correlation between the reversible CAC and EMP pathways. The genes coding for the two putative candidates for KOR, e.g., oxoacid synthase (oxoacid:ferredoxin oxidoreductase; oor gene), are organized in operons which comprise all four subunits (α, β, γ, and δ). Whereas for candidate 1 each subunit is encoded by one gene (operon oorC-D-A-B), the second candidate comprises subunits γ and δ in one ORF (operon oorCD-A-B). All genes overlap by 4 bp. The fifth operon comprises two genes coding for α and β subunits of succinyl-CoA synthetase (succ-D operon) which overlap by 4 bp. The sixth operon harbors four genes coding for the putative succinate dehydrogenase (sdh gene; candidate 1 Sdh/Frd). Whereas the downstream sdhA-located genes coding for the β, γ, and δ subunits overlap by 4 bp, the upstream-located sdhA gene coding for the α subunit is separated by 2 bp. The presence of putative promoter structures in front of sdhA only (Fig. 3B) indicates that all four genes form the sdhA-B-C-D operon. A second candidate, which encodes a putative fumarate reductase and comprises two genes with homology to sdhA/frdA and sdhB/frdB, was identified in the T. tenax genome. Here both ORFs overlap by 56 bp but are oriented in opposite directions, and the sdhB/frdB gene homolog overlaps by 4 bp with the adh gene which codes for a putative Zn-dependent alcohol dehydrogenase (adh-frdB operon).

(v) Glycogen metabolism. (a) Assignment of genes involved in glycogen metabolism. By the use of the T. tenax genome, gene homologs for almost all enzymes necessary for catalyzing
the synthesis and degradation of glycogen were found (Table 5). As determined on the basis of sequence similarity, seven genes are thought to encode sugar phosphate nucleotidyl transferases (snt) catalyzing the synthesis of the nucleotide-activated glucose. A possible role in glycogen synthesis is proposed for candidate 3 (snt3), which is found closely neighboring genes involved in glycogen metabolism (Fig. 3A). Because of the relatedness and numerous different annotations of gene homologs in the databases, a prediction of substrate specificity was not possible. Only for candidate 7, glucose-1-phosphate thymidylyltransferase (dTDP-glucose pyrophosphorylase [RfbA]; EC 2.7.7.24), can activity be suggested due to the closely neighboring genes coding for dTDP-glucose-4,6-dehydratase (RfbB; EC 4.2.1.46), dTDP-4-dehydroharnamose reductase (RfbD; EC 1.1.1.133), and dTDP-4-dehydroharnamose-3,5-epimerase (RfbC; EC 5.1.3.13). All four genes are involved in the biosynthesis of dTDP-harnamose, which represents a precursor for the O-specific polysaccharide chain in the lipopolysaccharide of the outer membrane in most gram-negative bacteria (36), thus raising questions about their functions in archaea.

Only one gene was identified which shows convincing similarity to genes encoding glycogen synthase, which catalyzes the formation of α-1-6 bonds. For glycogen degradation, four different genes were identified by sequence similarity: the
genes coding for glycogen phosphorylase, α-amylase, GAA, and α-glucosidase (maltase).

The glgP gene coding for glycogen phosphorylase was identified in T. tenax due to its similarity to genes encoding archael and bacterial glucan, glycogen, and maltodextrin phosphorylases. The glgP gene (1,458 bp), which codes for a protein of 485 amino acids, was expressed in E. coli. The recombinant protein was purified by heat precipitation and ion exchange chromatography and revealed a subunit molecular mass of 55 kDa (as determined by SDS-PAGE), a result in good agreement with the calculated molecular mass of 55.56 kDa. The enzyme catalyzes the Pi-dependent degradation of glycogen, forming glucose 1-phosphate. Glycogen phosphorylase follows Michaelis-Menten kinetics, with $K_m$ and $V_{max}$ values of 0.09 mg/ml and 0.9 U/mg of protein for glycogen and 0.09 mM and 0.9 U/mg of protein for P$_i$ at 50°C. Thus, the glycogen phosphorylase seems to play a major role in the phosphorolytic degradation of glycogen.

Genes encoding debranching enzymes (such as pullulanases) seem to be missing from the T. tenax genome. Since most of the GAAs described are able to hydrolyze 1,6-α-bonds also, however, this enzyme might compensate for the missing activity of debranching enzyme. The gaa gene is found closely neighboring the genes coding for the key enzymes of the modified ED pathways, indicating the close functional relationship between hydrolytic degradation of polysaccharides and the ED variants. Finally, one gene homolog was found with similarity to genes encoding phosphomannomutase or PGM. Whether the respective gene product exhibits PGM activity and thus is engaged in the catabolic and anabolic directions remains to be analyzed.

In summary, most gene homologs necessary for glycogen metabolism were identified in the T. tenax genome, and the glycogen phosphorylase, which is involved in the phosphorolytic degradation of glycogen, is characterized. The only missing counterparts are enzymes that are involved in the formation (branching enzyme) and hydrolysis of 1,6-α bonds. Whether these enzymes show no similarity to known enzymes or are generally absent and their activity is taken over by glycogen synthase and GAA remains to be proven.

(3) Organization of genes involved in glycogen metabolism.

Three of the seven gene homologs coding for sugar phosphate nucleotidyl transferases are organized in operons (Fig. 3; Table 5). The gene of candidate 1 (snt1) forms an operon with the gltX gene (AJ621274 and AJ621275; coding for glutamyl tRNA synthetase) and orfV (AJ621276; coding for an hypothetical protein). snt1 and gltX overlap by 14 bp, and gltX and orfV overlap by 4 bp, forming the snt1-gltX-orfV operon. The gene of candidate 2 (snt2) overlapps by 4 bp with orfZ (AJ621232), which encodes a hypothetical protein in P. aerophilum (orfZ-snt2 operon.) A functional connection is proposed for the putative glucose-1-phosphate thymidylyltransferase gene (rfbA; candidate 7) which forms an operon with the rfbB (4-bp overlap; dTDP-glucose-4,6-dehydratase; AJ621291), rfbD (10-bp overlap; dTDP-4-dehydrorhamnose reductase; AJ621290) and rfbC (3-bp gap; dTDP-4-dehydrorhamnose-3,5-epimerase; AJ621289) genes (rfbA-B-D-C operon). One of the genes involved in the glycogen metabolism, the genes encoding glycogen phosphorylase (glgP), α-amylase (amyA), glycogen synthase (glgA), and sugar phosphate nucleotidyl transferases (candidate 3; snt3), are clustered in the genome of T. tenax. The glgA gene is located upstream of a gene homolog (mthfs; AJ621295) which encodes a putative 5-formyltetrahydrofolate cyclo-igase and is oriented in the opposite direction. Both ORFs overlap by 22 bp. The amyA, glgA, and (separated by 44 bp) snt3 genes are
localization of the mitochondrial genes (75 bp) and in the same orientation as \( \text{glpP} \). Since both \( \text{amyA} \) and \( \text{glpA} \) overlap by 1 bp (\( \text{amyA-glpA} \) operon), we assume that a functional linkage existed between these two genes at least, suggesting that the \( \alpha \)-amylase is involved in glycogen metabolism. A similar combination of genes involved in glycogen degradation and synthesis has been previously reported for bacterial species (41).

As described above, the \( \text{gaa} \) gene coding for GAA is organized in an operon with genes coding for the key enzymes of the ED pathway (see "Organization of the ED genes," above). (vi) Trehalose metabolism. (a) Assignment of genes involved in trehalose metabolism. Sequence searches clearly indicated that trehalase is synthesized in \( T. \text{tenax} \) via the \( \text{otsA-otsB} \) pathway (Table 6). In this pathway, trehalase 6-phosphate is formed from glucose 6-phosphate and UDP-glucose by trehalase-6-phosphate synthase (TPS, or \( \text{otsA} \)) and trehalase 6-phosphate is dephosphorylated by trehalase-6-phosphate phosphatase (TPP, or \( \text{otsB} \)), as shown for \( \text{E. coli} \) and \( \text{S. cerevisiae} \) (16, 24). Strikingly, only one gene that was obviously a chimera composed of two domains, one coding for TPS and the other coding for TPP, was identified in \( T. \text{tenax} \). Thus, the protein represents a TPS-phosphatase (TPSP). First studies of the recombinant TPSP demonstrate that the encoded enzyme exhibits both activities (M. Zaparty, unpublished data). Besides this key enzyme of trehalose synthesis, further enzymes involved in the trehalose metabolism were identified: PGM and (NTP) sugar phosphate nucleotidylyltransferases, which are engaged in the glycogen metabolism, are also constituents of trehalose metabolism. Additionally, a gene homolog with similarity to either the gene encoding trehalase synthase (TreS; \( \text{treS} \) gene) or the gene encoding \( \alpha \)-trehalose phosphorylase (TreP; \( \text{treP} \) gene) seems to be present in \( T. \text{tenax} \). Whereas TreS catalyzes the conversion of the \( \alpha,1,1 \) linkage in maltose to \( \alpha,1,1 \) generating trehalose (specific for some bacteria, e.g., \( \text{Thermus aquaticus} \) and \( \text{Pimelobacter} \) (65), TreP catalyzes the phosphorolytic degradation of trehalose, forming glucose and glucose 1-phosphate. So far, no biochemical information is available concerning whether the gene product represents an alternative pathway for trehalose synthesis or is involved in trehalose degradation.

The bacterial and eucaryal OtsA-OtsB pathway has now been found in archaea, and the pathway in \( T. \text{tenax} \) is characterized by a bifunctional TPSP. So far, only the presence of the TreY-TreZ pathway has been described for archaea (e.g., \( \text{some} \) members of \( \text{Sulfolobales} \)), a pathway which is also present in some bacteria (e.g., \( \text{Arthrobacter} \) and \( \text{Micrococcus} \) spp.) (10, 30, 38). In this pathway, maltooligosyltrehalose synthase (TreY) converts the terminal \( \alpha,1,4 \)-linked residue of a glucose polymer in an \( \alpha,1,1 \) linkage and trehalose hydrolase (TreZ) cleaves the disaccharide off. Gene homologs of \( \text{treY} \) and \( \text{treZ} \) gene were not identified in the \( T. \text{tenax} \) genome, indicating that this pathway is absent. To learn whether the third pathway for trehalose synthesis (the \( \text{treS} \) pathway) is present, we demand enzymatic characterization of the \( \text{treS} \) gene product.

(b) Organization of genes involved in trehalose metabolism. The organization of genes also involved in glycogen metabolism is discussed above (see also Table 5). The \( \text{tpsp} \) gene coding for the unusual TPSP is localized downstream of an ORF coding for a putative glycosyltransferase (\( \text{gt} \)); the latter overlaps by 4 bp with an upstream-located ORF (\( \text{hp} \); AJ621288) which shows homology only to a gene coding for a hypothetical protein in \( \text{P. aerophilum} \) (Fig. 3; Table 6). Thus, at least the genes coding for the putative glycosyltransferase and the hypothetical protein seem to form an operon (\( \text{hp-gt} \) operon). Also, an ORF (\( \text{orfY; AJ621342} \)) (upstream of the \( \text{treS-treP} \) gene coding for the trehalose synthase or trehalose phosphorylase) was found with similarity to a gene encoding a hypothetical protein in \( \text{P. aerophilum} \), \( \text{S. sofataarius} \), and \( \text{S. tokodaii} \). Both genes overlap by 4 bp and form the \( \text{orfY-treS-treP} \) operon (Fig. 3), but so far we have no indication about the function of the two genes.

TABLE 6. Identified genes coding for enzymes of the trehalose metabolism in \( T. \text{tenax} \)

<table>
<thead>
<tr>
<th>Enzymea</th>
<th>Candidate no.</th>
<th>EC no.</th>
<th>COG no.b</th>
<th>Gene</th>
<th>Accession no.</th>
<th>Evidence value and best hitc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phosphoglucomutase</td>
<td>1</td>
<td>2.7.5.1</td>
<td>1109</td>
<td>\text{pgm}</td>
<td>AJ621332</td>
<td>1.0e-154 ( \text{P. aerophilum} ), phosphomannomutase</td>
</tr>
<tr>
<td>Sugar phosphate nucleotidylyl transferase (7 candidates)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.7.7.4</td>
<td>1208</td>
<td>\text{snt}</td>
<td>AJ621273</td>
<td>1.9e-91 ( \text{P. aerophilum} ), sugar phosphate nucleotidylyl transferase</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trehalose-6-phosphate synthase-phosphatase</td>
<td>0380</td>
<td>1877</td>
<td>\text{tsp}</td>
<td>AJ621287</td>
<td>0.0 ( \text{P. aerophilum} ), trehalose-6-phosphate synthase</td>
<td></td>
</tr>
<tr>
<td>Trehalase synthase/Trehalose phosphorylase</td>
<td>5.4.99.16</td>
<td>1877</td>
<td>\text{treS/treP}</td>
<td>AJ621341</td>
<td>1.0e-176 ( \text{P. aerophilum} ), trehalose synthase</td>
<td></td>
</tr>
</tbody>
</table>

---

\( ^a \) The enzymes involved in trehalose degradation are underlined, and those involved in its biosynthesis are in bold.

\( ^b \) For definition, see Table 1, footnote \( b \).

\( ^c \) For definition, see Table 1, footnote \( c \). In cases in which two or more candidates were identified in the genome, the respective evidence values are given.

\( ^d \) Final EC number not yet assigned.
Conclusions. T. tenax genome data reveal the presence of pathways involved in central carbohydrate metabolism as well as in glycogen and trehalose turnover (Fig. 4). So far, T. tenax and S. solfataricus are the only facultative heterotrophic archaea sequenced. Thus, the presented analysis not only contributes to the diversity of carbon metabolism in archaea but also represents an important step towards understanding of archaeal metabolic networks, which allow adaptation to alternative growth conditions.

The identification of genes coding for HK, PGI, PPDK, GDH, GAD, KD(P)GA, KDGK, (glycogen) phosphorylase, and TPSP was confirmed by the functions of their recombinant gene products. In some cases, operon organization provides information about the function of genes, facilitating the reconstruction of pathways. Several genes organized in operons seem to overlap by 4 bp (the coding regions overlap by 1 bp) (Fig. 3), and coordinated transcription has been shown for three overlapping bigenic operons (orfX-hxk, fba-pfp, and pgk-gap) (5, 11, 54). For the acn-tpi operon, where the genes are separated by 40 bp, cotranscription was demonstrated by Northern analysis (63). Putative TATA boxes and B recognition element (BRE) sites were found only in front of the first gene of an operon (3, 59, 60) (except for the acn-tpi and rfbA-B-D-C operon), whereas putative Shine-Dalgarno sequences were often absent (29) (Fig. 3B). These results, as well as those determined for the transcription start sites of the single PK gene (pyk) (48) and the fbp-pfp operon (54), suggest also that in T. tenax, the translation of single genes and genes that are first in operons proceeds via leaderless transcripts (58, 64). The occurrence of monocistronic messages of the fba, pfp, pgk, gap, acn, and tpi genes, as well as the agreement of the N-terminal sequence of the PPi-PFK purified from T. tenax with the predicted translation start, suggests that posttranscriptional processing occurs (5, 11, 54).

The genome analysis strongly suggests the presence of three pathways for carbohydrate metabolism: (i) the EMP variant and (ii) the semiphosphorylative and (iii) the nonphosphorylative versions of the ED pathway. At the moment, we can only speculate about the physiological significance of the three different pathways. The organization of ED genes suggests that the ED variants are involved in hydrolytic degradation of glycogen, whereas the EMP variant is responsible for the phosphorolytic breakdown. The energy demand of the cell seems to have a strong influence on the selection of the different pathways. Whereas no ATP is generated by the ED variants, one ATP is gained using the EMP variant (taking into account that PPi, the phosphoryl donor of the phosphofructokinase, is a waste product of the cell). In T. tenax, GAP, the common intermediate of EMP pathway and semiphosphorylative ED version, can be converted into 3-phosphoglycerate in a one-step reaction—via GAPN or GAPOR—thus circumventing the formation of the extremely heat-labile intermediate 1,3-bisphosphoglycerate. Since the nonphosphorylative ED pathway avoids the most heat-labile intermediates (1,3-bisphosphoglycerate, GAP, and dihydroxyacetone phosphate), this pathway could be preferred at the upper temperature range of growth.

Further, the sequence data support the idea of the existence of a reversible CAC in T. tenax, indicating that—as already suggested—the organism uses the reaction cycle not only for catabolic purposes but also for CO₂ fixation (1). Almost all

![FIG. 4. The central carbohydrate metabolism in T. tenax. A selection of growth substrates is given at the top; the labels of the different pathways are marked by grey-shaded boxes.](http://jb.asm.org/)
gene homologs necessary for the glycogen metabolism (with the exception of the branching and debranching enzyme) were identified in the \textit{T. tenax} genome. In addition, trehalose synthesis was shown to proceed via the OtsA-OtsB pathway in \textit{T. tenax}. No evidence was found for a functional PPP, which has been thought to be essential for generation of pentoses and NADPH for anabolic purposes. Together with previous genome analyses of methanooarchaea and hyperthermophilic archaea, this finding suggests that the oxidative PPP is generally absent from archaea or that an analogous pathway exists with enzymes showing no significant similarity to known bacterial and eucaryal counterparts.


central carbohydrate metabolism of \textit{T. tenax}